The diagnostic value of MR Arthrography assessment, in patients with traumatic anterior shoulder instability

Susan van Grinsven

The diagnostic value of MR Arthrography assessment, in patients with traumatic anterior shoulder instability

Susan van Grinsven

The publication of this thesis was financially supported by
Stichting Rijnstate Vriendenfonds
Anna Fonds, Leiden
ABN AMRO Bank N.V.
Bauerfeind Benelux B.V.

Photo cover (with acknowledgment to model Daphne) by
Ian Beck, Apeldoorn, info@ianbeck.nl
Printed and bound by
Ipskamp printing, Enschede, info@ipskampprinting.nl

ISBN
978-94-028-0114-9

The diagnostic value of MR Arthrography assessment, in patients with traumatic anterior shoulder instability

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus,
volgens besluit van het college van decanen
in het openbaar te verdedigen op
woensdag 25 mei 2016
om 12.30 uur precies

door

Susan van Grinsven
geboren op 19 juli 1968
te Boxmeer

Promotor
Prof. dr. A. van Kampen

Copromotor
Dr. C.J.M. van Loon (Rijnstate Arnhem)

Manuscriptcommissie
Prof. dr. W.M. Prokop (voorzitter)
Prof. dr. W.J.J. Assendelft
Prof. dr. G.M.M.J. Kerkhoffs (Universiteit van Amsterdam)

Paranimfen
Karin van Leeuwen - van Grinsven
Marij Lamers

Voor mijn moeder

Contents

Chapter 1	Introduction and aims	9
Chapter 2	Klinische les: Anterieure schouderluxaties, een behandeling op maat	19
Chapter 3	MR arthrography of traumatic anterior shoulder lesions showed modest reproducibility and accuracy when evaluated under clinical circumstances	31
Chapter 4	Does the experience level of the radiologist, assessment in consensus or the addition of the ABER position view improve the diagnostic reproducibility and accuracy of MR arthrography of the shoulder	45
Chapter 5	Reproducibility and accuracy of instability related shoulder MRA's: Orthopaedic surgeons outperform radiologists	71
Chapter 6	Are radiologists superior to orthopaedic surgeons in diagnosing instability related shoulder lesions on magnetic resonance arthrography? A multicentre reproducibility and accuracy study	95
Chapter 7	A feedback protocol improves the diagnostic performance of MR Arthrography by experienced musculoskeletal radiologists in patients with traumatic anterior shoulder instability	117
Chapter 8	Summary and conclusions	137
Chapter 9	Samenvatting en conclusies	147
List of Publications		159
Curriculum Vitae		161

Chapter | 1

Introduction and aims

Introduction

The glenohumeral joint is the most mobile joint in the human body. Its stability depends on static and dynamic stabilizers. Static stabilization is provided by the glenoid concavity, negative intra-articular pressure of the glenohumeral joint, labral height and glenohumeral ligaments. Dynamic stabilization is achieved through proprioceptive and neuromuscular control of the rotator cuff, biceps, deltoid, triceps and scapulothoracic musculature. To a certain extend the dynamic and static stabilizers can compensate each others insufficiency. Only if the compensation mechanism fails shoulder dislocation can occur. [1-2]

In the Netherlands the yearly incidence of shoulder dislocation is estimated at 38 per 100.000 persons. 85% of the shoulder joint dislocations are in anterior direction and 96% are of traumatic origin after forced abduction / exorotation of the shoulder. After this damaging manoeuvre anterior lesions such as bony Bankart, Bankart, joint capsule or Hill-Sachs lesions are usually seen in younger patients while rotator cuff tears or greater humeral tuberosity fractures are more often observed in older patients. [3-6]

When confronted with a young (age 15-29) and active patient with instability complaints after reduced traumatic anterior shoulder dislocation (TASI) the general practitioner should refer the patient to an orthopaedic shoulder surgeon to prevent further damage, as the recurrence rate of TASI is up to 95%. [2,7]

Although instability related shoulder lesions can be strongly suspected by patient history, physical examination and standard x-rays alone [3,8-13] additional diagnostic imaging is used by the orthopaedic surgeon to support the clinical findings. Confirmation of lesion type, lesion location and lesion severity can guide the treatment decision from nonsurgical to arthroscopic or even open surgical approach. [4,8,10,14-22]

Currently magnetic resonance arthrography (MRA) is considered the “pre-arthroscopy” diagnostic reference standard for patients with suspected TASI. [4,8,10-11, 16,18,20-21,23-35] Its utility depends on its high accuracy and high reproducibility. Accuracy is the ability of a test to discriminate between conditions of interest. Reproducibility is the ability to achieve measurement repetition of a variable. [3,6,10,12,15-16,19-22,29,33-34,37-44]

In our daily clinical practice, however, disagreement between radiologists and orthopaedic surgeons about the presence of instability related lesions on MRA frequently occurred. This is a reproducibility problem. Moreover, after stabilizing surgery discrepancies were noticed between the radiologic MRA results and surgical outcomes. This is an accuracy problem.

As a high diagnostic performance is conditional to proper patient selection and successful treatment [4,8,10,14-22] we wondered whether the diagnostic MRA performance we reached in our daily clinical practice approximates the high end of the range mentioned in previous literature: kappa = 0.61-0.80, accuracy = 79%-100% [3,6,10,12,15-16,19-22,29,33-34,37-44] and if not, what the options were to improve the situation.

Reviewing the literature we decided, that the best way to proceed was to evaluate the influence of observer experience, assessment in consensus and the additional value of the abduction external rotation view (ABER) on the diagnostic performance of TASI-related MRA. [5, 8,10,12,22-24,26,28-29,31-32,35,39-41,45-48] As radiologists can learn from one another by discussing shoulder lesions, that are difficult to diagnose at MRA, it is suggested, that consensus instead of individual assessment by experienced musculoskeletal radiologists improves the reproducibility and accuracy of MRA. [5,8,24,29,39,46,48] The ABER view is presumed to optimize the diagnostic performance of MRA by improving the visualization of rotator cuff partial thickness tears and any present lesion in the anterior labroligamentous complex through distraction and increased penetration of contrast material into a tear. [10,12,22-23,26,28,32,35,41,45-47]

Although we hypothesed, that the consensus assessment of TASI-related MRA's with ABER view would optimize the results of our experienced musculoskeletal radiologists, we expected, that there would always be cases left where the clinical diagnosis is not confirmed by MRA and disagreement exists between the radiologist and orthopaedic surgeon about the presence of instability related shoulder lesions on MRA. [13,33-34,37,41] To help the orthopaedic surgeon to decide which MRA interpretation is likely to be the most accurate we decided to compare the diagnostic MRA performance of the two disciplines in order to plan a successful treatment strategy.

Motivated by the results of our studies we ultimately sought alternative ways to optimize our diagnostic process. We reasoned, that we should be able to improve the agreement between our radiologists and orthopaedic surgeons and the accuracy of our radiologists through systematic personal feedback of the MRA assessment after stabilizing surgery, as it would enable our radiologists and orthopaedic surgeons to discuss discrepancies and fine-tune their agreement about lesion definition interpretation. [3,9,41]

Research questions addressed in this thesis

1. At what point should a general practitioner (GP) refer a patient with anterior shoulder dislocation or complaints after reduced anterior shoulder dislocation to an orthopaedic shoulder surgeon to avoid further damage / complaints?
2. What is the reproducibility and accuracy of an 1.5 Tesla magnetic resonance arthrography (MRA) of patients with traumatic anterior shoulder instability (TASI) under conditions resembling the daily clinical practice of radiologists?
3. What is the influence of the experience level of the interpreting radiologist on the diagnostic reproducibility and accuracy of an 1.5 Tesla MRA of patients with TASI?
4. What is the influence of assessment in consensus compared to individual assessment of radiologists on the diagnostic performance of an 1.5 Tesla MRA of patients with TASI?
5. What is the additional value of the abduction and external rotation (ABER) sequence on the diagnostic performance of an 1.5 Tesla MRA of patients with TASI?

6. Is the diagnostic reproducibility and accuracy of experienced musculoskeletal radiologists assessing 1.5 Tesla MRA's of patients with TASI superior to experienced orthopaedic shoulder surgeons?
7. Does a feedback protocol improve the diagnostic reproducibility and accuracy of experienced musculoskeletal radiologists assessing 1.5 Tesla MRA's of patients with TASI?

Outline and aims of this thesis

In **chapter 2** we present a clinical lesson about anterior shoulder dislocation for GP's. We address the static and dynamic stabilizers of the shoulder, anterior shoulder dislocation, anterior shoulder instability and the typical workup of a GP when presented with a patient with anterior shoulder dislocation or complaints of anterior shoulder instability. The workup of a GP normally includes patient history, physical examination, diagnosis and management strategies. In the same chapter indications for a GP to refer the patient to a shoulder specialist and the evidence based workup of the orthopaedic shoulder surgeon are outlined. This orthopaedic workup includes patient history, physical examination, diagnostic tests such as MRA, diagnosis, conservative or operative management strategies and referral for rehabilitation if indicated.

Chapter 3 describes the results of a retrospective diagnostic MRA study. In view of the usefulness of MRA to select subjects eligible for stabilizing surgery we examine the reproducibility and accuracy of an 1.5 Tesla MRA of patients with traumatic anterior shoulder instability under conditions resembling daily clinical practice. These conditions include time constraints, incidental suboptimal MRA quality, variation in observer's experience level and independent assessment. 61 MRA's are independently assessed by 2 radiologists with different experience levels for the presence of a cuff lesion, Hill-Sachs lesion, bony Bankart lesion, Bankart lesion, greater humeral tuberosity fracture, SLAP lesion and joint capsule lesion. Suboptimal MRA's due to time constraints are not excluded.

Systematic accurate MRA diagnosis of each separate radiologist is achievable through high reproducibility between radiologists. MRA reproducibility and accuracy rates are, however, highly variable in the present literature. As a possible explanation a suboptimal experience level of the interpreting radiologist, individual instead of consensus assessment and the omission of the ABER view are mentioned. In order to improve the diagnostic performance of radiologists we, therefore, examine the influence of observer experience, assessment in consensus and the added value of the ABER view on the diagnostic reproducibility and accuracy of an 1.5 Tesla MRA of patients with TASI in **chapter 4.** In this prospective diagnostic MRA study 6 radiologists (independent assessment) and 3 teams of radiologists (consensus assessment) with different experience levels assess 58 shoulder MRA's for 7 different lesion types with and without the ABER view.

In our daily clinical practice disagreement between radiologists and orthopaedic surgeons about the presence of instability related shoulder lesions on MRA frequently occurs. Consequently, the orthopaedic surgeon has to decide which MRA interpretation is likely to be the most accurate in order to plan a successful treatment strategy. In **chapter 5** we prospectively compare the diagnostic reproducibility and accuracy of experienced musculoskeletal radiologists with those of experienced orthopaedic shoulder surgeons interpreting an 1.5 Tesla MRA of patients with TASI. For this purpose 2 radiologists and 2 orthopaedic surgeons assess 7 different instability related shoulder lesions on 58 MRA's. The 7 lesions comprise greater humeral tuberosity fractures, cuff, Hill-Sachs, bony Bankart, Bankart, SLAP and joint capsule lesions.

Accurate pre-operative diagnosis leads to better treatment decisions and will ultimately result in better healthcare. In **chapter 6** we compare the diagnostic reproducibility and accuracy of musculoskeletal radiologists with those of orthopaedic shoulder surgeons in two large medical centers assessing an 1.5 Tesla MRA of patients with TASI to verify the rather surprising results of chapter 5. In this prospective multi center diagnostic MRA study 4 radiologists, 4 orthopaedic surgeons, 2 radiologic and 2 orthopaedic teams of 2 different medical centers assess the same 7 instability related shoulder lesions on 45 surgically confirmed MRA's.

To improve suboptimal reproducibility and accuracy rates we developed a feedback protocol in which radiologist systematically receive personal feedback of their MRA assessment after surgery, enabling radiologists and orthopaedic surgeons to discuss discrepancies and fine-tune their agreement about lesion definition interpretation. In **chapter 7** we evaluate the diagnostic performance of an 1.5 Tesla MRA by experienced musculoskeletal radiologists of patients with TASI after feedback protocol execution. For this purpose 45 old surgically confirmed MRA's are used to enhance personal feedback, discuss differences and fine-tune agreement between the 2 disciplines in a short period of time. After which 2 experienced musculoskeletal radiologists are asked to assess the 7 instability related lesions on 20 new surgically confirmed MRA's to test whether the diagnostic performance has indeed improved.

Finally, **chapter 8 and chapter 9** summarise the studies described in this theses in English and in Dutch. Conclusions are drawn and the implications for future research and daily clinical practice of radiologists and orthopaedic shoulder surgeons are outlined.

References

1. Gaskill TR, Taylor DC, Millett PJ. Management of multidirectional instability of the shoulder. J Am Acad Orthop Surg. 2011;19:758-767.
2. Sileo MJ, Joseph S, Nelson CO, Botts JD, Penna J. Management of acute glenohumeral dislocations. Am J Orthop. 2009;38:282-290.
3. van Kampen DA, van den Berg T, van der Woude HJ, Castelein RM, Terwee CB, Willems WJ. Diagnostic value of patient characteristics, history, and six clinical tests for traumatic anterior shoulder instability. J Shoulder Elbow Surg 2013;22:1310-19.
4. Acid S, Le Corroller T, Aswad R, Pauly V, Champsaur P. Preoperative imaging of anterior shoulder instability: diagnostic effectiveness of MDCT arthrography and comparison with MR arthrography and arthroscopy. AJR Am J Roentgenol 2012;198(3):661-667.
5. van der Veen HC, Collins JP, Rijk PC. Value of magnetic resonance arthrography in post-traumatic anterior shoulder instability prior to arthroscopy: a prospective evaluation of MRA versus arthroscopy. Arch Orthop Trauma Surg 2012;132(3):371-375.
6. Matsen FA, Rockwood CA. Glenohumeral instability. The Shoulder. Matsen FA, Rockwood CA, editors. WB Saunder Company, Philadelphia 1998;611-745.
7. Cutts S, Prempeh M, Drew S. Anterior shoulder dislocation. Ann R Coll Surg Engl. 2009;91:2-7.
8. van Grinsven S, Kesselring FO, van Wassenaer-van Hall HN, Lindeboom R, Lucas C, van Loon CJ. MR arthrography of traumatic anterior shoulder lesions showed modest reproducibility and accuracy when evaluated under clinical circumstances. Arch Orthop Trauma Surg 2007;127(1):11-17.
9. van Grinsven S, Hagenmaier F, van Loon CJ, van Gorp MJ, van Kints MJ, van Kampen A. Does the experience level of the radiologist, assessment in consensus, or the addition of the abduction and external rotation view improve the diagnostic reproducibility and accuracy of MRA of the shoulder?. Clinical Radiology 2014; 69:1157-1164.
10. Smith TO, Drew BT, Toms AP. A meta-analysis of the diagnostic test accuracy of MRA and MRI for the detection of glenoid labral injury. Arch Orthop Trauma Surg 2012;132(7):905-919.
11. Bak K, Wiesler ER, Poehling GG. Consensus statement on shoulder instability. Arthroscopy 2010;26(2):249-255.
12. Oh JH, Kim JY, Choi JA, Kim WS. Effectiveness of multidetector computed tomography arthrography for the diagnosis of shoulder pathology: comparison with magnetic resonance imaging with arthroscopic correlation. J Shoulder Elbow Surg 2010;19(1):14-20.
13. Pandya NK, Colton A, Webner D, Sennett B, Huffman GR. Physical examination and magnetic resonance imaging in the diagnosis of superior labrum anterior-posterior lesions of the shoulder: a sensitivity analysis. Arthroscopy 2008;24(3):311-317.
14. Bencardino J, Gyftopoulos S, Palmer W. Imaging in Anterior Glenohumeral Instability. Radiology 2013;269(2):323-337.
15. Khedr SA, Kassem HM, Azab MA. Comparative study of direct MR arthrography and CT arthrography with arthroscopic correlation in preoperative evaluation of anterior shoulder instability. Egypt J Radiol Nucl Med 2013;44:817-827.
16. Simão MN, Nogueira-Barbosa MH, Muglia VF, Barbieri CH. Anterior shoulder instability: correlation between magnetic resonance arthrography, ultrasound arthrography and intraoperative findings. Ultrasound Med Biol 2012;38:551-560.

17. van der Woude HJ, Vanhoenacker FM. MR Arthrography in Glenohumeral Instability. JBR–BTR 2007; 90:377-383.
18. Amin MF, Youssef AO. The diagnostic value of magnetic resonance arthrography of the shoulder in detection and grading of SLAP lesions: Comparison with arthroscopic findings. Eur J Radiol 2012;81(9): 2343-2347.
19. Genovese E, Spano E, Castagna A, Leonardi A, Angeretti MG, Callegari L, et al. MR-arthrography in superior instability of the shoulder: correlation with arthroscopy. Radiol med 2013; 118:1022–1033.
20. Iqbal HJ, Rani S, Mahmood A, Brownson P, Aniq H. Diagnostic value of MR arthrogram in SLAP lesions of the shoulder. Surgeon 2010;8(6):303-309.
21. de Jesus JO, Parker L, Frangos AJ, Nazarian LN. Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. AJR Am J Roentgenol 2009;192(6):1701-1707.
22. Tian CY, Cui GQ, Zheng ZZ, Ren AH. The added value of ABER position for the detection and classification of anteroinferior labroligamentous lesions in MR arthrography of the shoulder. Eur J Radiol 2013; 82(4):651-657.
23. Choi JA, Suh SI, Kim BH, Cha SH, Kim MG, Lee KY, et al. Comparison between conventional MR arthrography and abduction and external rotation MR arthrography in revealing tears of the antero-inferior glenoid labrum. Korean J Radiol 2001;2:216-221.
24. Jonas SC, Walton MJ, Sarangi PP. Is MRA an unnecessary expense in the management of a clinically unstable shoulder? Acta Orthop 2012;83:1-4.
25. Sanders TG, Morrison WB, Miller MD. Imaging techniques for the evaluation of glenohumeral instability. Am J Sports Med 2000;28:414-434.
26. Ito Y, Sakai T, Tomo H, Nakao Y, Inui K, Koike T, et al. Computerized assessment of Bankart lesions under tension with magnetic resonance arthrography. J Shoulder Elbow Surg 2005;14:247-251.
27. Chandnani VP, Yeager TD, DeBerardino T, Christensen K, Gagliardi JA, Heitz DR, et al. Glenoid labral tears: prospective evaluation with MRI imaging, MR arthrography, and CT arthrography. AJR Am J Roentgenol 1993;161:1229-1235.
28. Cochet H, Couderc S, Pele E, Amoretti N, Moreau-Durieux MH, Hauger O, et al. Rotator cuff tears: should abduction and external rotation (ABER) positioning be performed before image acquisition? A CT arthrography study. Eur Radiol 2010;20:1234-1241.
29. Holzapfel K, Waldt S, Bruegel M, Paul J, Heinrich P, Imhoff AB, et al. Inter- and intraobserver variability of MR arthrography in the detection and classification of superior labral anterior posterior (SLAP) lesions: evaluation in 78 cases with arthroscopic correlation. Eur Radiol 2010;20:666-673.
30. Lee MJ, Motamedi K, Chow K, Seeger LL. Gradient-recalled echo sequences in direct shoulder MR arthrography for evaluating the labrum. Skeletal Radiol 2008;37:19-25.
31. Lee SY, Lee JK. Horizontal component of partial-thickness tears of rotator cuff: imaging characteristics and comparison of ABER view with oblique coronal view at MR arthrography initial results. Radiology 2002;224:470-476.
32. Matava MJ, Purcell DB, Rudzki JR. Partial-thickness rotator cuff tears. Am J Sports Med 2005;33:1405-1417.
33. Schreinemachers SA, van der Hulst VP, Willems WJ, Bipat S, van der Woude HJ. Detection of partial-thickness supraspinatus tendon tears: is a single direct MR arthrography series in ABER

position as accurate as conventional MR arthrography? Skeletal Radiol 2009;38:967-975.
34. Schreinemachers SA, van der Hulst VP, Jaap Willems W, Bipat S, van der Woude HJ. Is a single direct MR arthrography series in ABER position as accurate in detecting anteroinferior labroligamentous lesions as conventional MR arthography? Skeletal Radiol 2009;38:675-683.
35. Song HT, Huh YM, Kim S, Lee SA, Kim SJ, Shin KH, et al. anterior-inferior labral lesions of recurrent shoulder dislocation evaluated by MR arthrography in an adduction internal rotation (ADIR) position. J Magn Reson Imaging 2006;23:29-35.
36. Drescher, R, von Rothenburg T, Ludwig J, Köster O, Schmid G. Direct MR-arthrography of the shoulder with maximum capsular distension for surgical planning. Fortschr Röntgenstr 2004;176(10):1469-1474.
37. Dinter DJ, Martetschläger F, Büsing KA, Schönberg SO, Scharf HP, Lehmann LJ. Shoulder Injuries in Overhead Athletes: Utility of MR Arthrography. Sportverl Sportschad 2008;22:146–152.
38. Toyoda H, Ito Y, Tomo H, Nakao Y, Koike T, Takaoka K. Evaluation of rotator cuff tears with magnetic resonance arthrography. Clin Orthop Relat Res 2005;439:109-115.
39. Theodoropoulos JS, Andreisek G, Harvey EJ, Wolin P. Magnetic resonance imaging and magnetic resonance arthrography of the shoulder: dependence on the level of training of the performing radiologist for diagnostic accuracy. Skeletal Radiol 2010;39(7):661-667.
40. Jung JY, Jee WH, Chun HJ, Ahn MI, Kim YS. Magnetic resonance arthrography including ABER view in diagnosing partial-thickness tears of the rotator cuff: accuracy, and inter- and intra-observer agreements. Acta Radiol 2010;51(2):194-201.
41. Halma JJ, Eshuis R, Krebbers YM, Weits T, de Gast A. Interdisciplinary inter-observer agreement and accuracy of MR imaging of the shoulder with arthroscopic correlation. Arch Orthop Trauma Surg 2012;132(3):311-320.
42. Applegate GR, Hewitt M, Snyder SJ, Watson E, Kwak S, Resnick D. Chronic labral tears: value of magnetic resonance arthrography in evaluating the glenoid labrum and labral-bicipital complex. Arthroscopy 2004;20(9):959-963.
43. Hayes ML, Collins MS, Morgan JA, Wenger DE, Dahm DL. Efficacy of diagnostic magnetic resonance imaging for articular cartilage lesions of the glenohumeral joint in patients with instability. Skeletal Radiol 2010;39(12):1199-1204.
44 Jana M, Srivastava DN, Sharma R, Gamanagatti S, Nag HL, Mittal R, et al. Magnetic resonance arthrography for assessing severity of glenohumeral labroligamentous lesions. J Orthop Surg 2012;20(2):230-235.
45. Herold T, Bachthaler M, Hamer OW, Hente R, Feuerbach S, Fellner C, et al. Indirect MR arthrography of the shoulder: use of abduction and external rotation to detect full- and partial-thickness tears of the supraspinatus tendon. Radiology 2006;240:152-160.
46. Jee WH, McCauley TR, Katz LD, Matheny JM, Ruwe PA, Daigneault JP. Superior labral anterior posterior (SLAP) lesions of the glenoid labrum: reliability and accuracy of MR arthrography for diagnosis. Radiology 2001;218:127-132.
47. Chun KA, Kim MS, Kim YJ. Comparisons of the various partial-thickness rotator cuff tears on MR arthrography and arthroscopic correlation. Korean J Radiol 2010;11:528-535.
48. Reuss BL, Schwartzberg R, Zlatkin MB, Cooperman A, Dixon JR. Magnetic resonance imaging accuracy for the diagnosis of superior labrum anterior-inferior lesions in the community setting: eighty-three arthroscopically confirmed cases. J Shoulder Elbow Surg 2006;15:580-585.

Chapter | 2

Klinische les: Anterieure schouderluxaties, een behandeling op maat

Susan van Grinsven, Roy Marinus,
Corné van Loon, Andre Haverkort,
Albert van Kampen

Submitted

Samenvatting

Huisartsen worden regelmatig geconsulteerd t.a.v. schouderklachten. Hiertoe behoren incidenteel ook patiënten met (doorgemaakte) anterieure schouderluxatie. Instabiliteit door hyperlaxiteit kan gemakkelijk door de huisarts worden vastgesteld m.b.v. de sulcustest, anterior apprehensiontest en Beighton-schaal. Doorverwijzen naar een schouderspecialist is hier alleen nuttig als uitgebreide fysiotherapie niet helpt. Voor goede follow-up bij acute traumatische schouderluxatie is directe doorverwijzing naar de tweede lijn wel noodzakelijk. Ook uitblijvend herstel van "range-of-motion", totale afwezigheid van kracht of blijvende instabiliteit na (primaire of secundaire) schouderluxatie zijn indicaties voor de huisarts om snel door te verwijzen naar een specialist met kennis t.a.v. schouderpathologie. Na passende diagnostiek en repositie is vervolgbehandeling afhankelijk van luxatiereden, leeftijd, activiteitenniveau en begeleidend letsel. Naast keuze uit één van de verschillende operatieve ingrepen behoort ook conservatieve fysiotherapeutische behandeling tot de mogelijkheden. Behandeling van anterieure schouderluxatie (na repositie) is altijd gericht op het verminderen van pijnklachten, optimaliseren van functie en stabiliteit van het schoudergewricht en voorkomen van verdere schouderluxaties.

Inleiding

Het glenohumerale gewricht is het meest mobiele gewricht van het menselijk lichaam. Dit gewricht kent hierdoor ook de meeste luxaties, waarvan de anterieure vorm na geforceerde abductie/exorotatie het meeste voorkomt.

De schouder ontleent zijn stabiliteit aan zowel statische als dynamische structuren. Statische stabilisatoren zijn de glenoid concaviteit, negatieve intra-articulaire druk, labrumhoogte en glenohumerale ligamenten. Dynamisch stabiliteit wordt verkregen door proprioceptieve en neuromusculaire controle van de rotator-cuff, biceps, deltoideus, triceps en scapulothoracale spieren. De dynamische en statische stabilisatoren kunnen elkaars insufficiëntie compenseren. Als het compensatiemechanisme faalt ontstaat schouder-instabiliteit waardoor naast pijn en instabiel gevoel ook (re)luxaties kunnen optreden. [1,2]

Schouderinstabiliteit, het onvermogen om de schouder in de kom te houden, kan ontstaan door een trauma waarbij statische en/of dynamische structuren beschadigd raken. Daarnaast kunnen aangeboren systemische aandoeningen zoals m. Ehlers-Danlos en m. Marfan gewrichtshyperlaxiteit met schouderinstabiliteit veroorzaken. 4-13% van de mensen heeft gewrichtshypermobiliteit zonder onderliggende oorzaak. De grens tussen normale en hyperlaxiteit (overmatige glenohumerale beweeglijkheid) is echter arbitrair. [3,4]

Luxaties worden geclassificeerd naar de humerusstand t.o.v. het glenoid. De meest-voorkomende schouderluxatie is de anterieure luxatie (anterieur 85-98%, posterior 2-3%, inferior 0,5%). [2] De incidentie in de hele bevolking is 1,7%. Dit getal kan verdubbelen in een populatie met hoge fysieke belasting. Een luxatie kan schade van labrum, de rotator-cuff, de benige-, capsuloligamentaire- en neurovasculaire structuren veroorzaken.

Aanleiding voor deze klinische les

Patiënten met schouderklachten zullen zich wenden tot huisartsen, spoedeisende hulp, traumachirurgen en orthopedisch chirurgen. [5] Gemiddeld wordt een huisarts driemaal per week geconsulteerd. Echter, met een jaarlijkse incidentie van 32 schouderluxaties per 100.000 inwoners in Nederland zal een huisarts niet vaak geconsulteerd worden i.v.m. klachten van (of na) een luxatie. [6] De behandeling van anterieure schouderluxatie kan per patiënt sterk verschillen (afhankelijk van luxatie-oorzaak, leeftijd, activiteitenniveau en ernst begeleidend letsel) [7] en is meer uitgekristalliseerd dan 10 jaar geleden. Terwijl bij de ene patiënt diagnose en behandeling prima door de huisarts is vast te stellen is voor een andere patiënt vroege doorverwijzing naar een schouderspecialist onontbeerlijk.

Doelstelling

In deze klinische les willen we de huisarts een praktische handreiking bieden bij klachten van (of na) anterieure schouderluxatie. Wanneer is verwijzen nodig?

Patiënt A, B en C

Bij de huisarts meldt zich **patiënt A.** Zij is een 22-jarige vrouw met linkerschouderklachten sinds een jaar. Ze heeft 4 maanden geleden tijdens een badmintonsmash een anterieure schouder(sub)luxatie gehad, die ter plaatse is gereponeerd. Alhoewel het na kortdurende fysiotherapie even beter ging zijn de klachten hierna toenemend verergerd. De schoudervoorzijde voelt pijnlijk en instabiel tijdens sporten. Smashen durft ze niet meer. Ze wordt wakker wanneer ze met haar linkerhand boven haar hoofd slaapt en nu ze ook klachten krijgt op haar werk, bij het dragen of hoog in de kast zetten van zware dossiermappen, is de maat vol.

Bij het lichamelijk onderzoek wordt geen passieve bewegingsbeperking of abductiepijn gevonden. [6] Wel vallen “erg ruime” bewegingsuitslagen van de linkerschouder en forse ellebooghyperextensie op. Omdat de huisarts gezien leeftijd en geslacht aan schouderinstabiliteit door gewrichtshyperlaxiteit denkt wordt een familieanamnese afgenomen. [1,4,7] Deze geeft geen bijzonderheden. De patiënte wordt uitgebreid voorgelicht t.a.v. gevonden oorzaak, werkgerelateerde ergonomische oplossingen (zware mappen onder schouderhoogte in kast plaatsen), mogelijk langdurig klachtenbeloop en behandel(on) mogelijkheden. Bij schouderhypermobiliteit is operatie niet de eerste optie. Bij hypermobiliteit is rotator-cuff letsel of intra-articulaire schade niet waarschijnlijk en indien er capsulolabrale afwijkingen zijn dan is het resultaat zeer matig. [1,4,7] De huisarts adviseert daarom langdurige stabiliserende fysiotherapie met aansluitend onderhoudsprogramma (kernpunt 1). Het succespercentage is 83-88%. [1,4] De patiënte wil ondanks positieve ervaring met fysiotherapie echter eerst mogelijke schade uitsluiten. Gezien leeftijd, sportieve levensstijl, doorgemaakte schouderluxatie en problemen bij werk en sport besluit de huisarts haar toch maar voor verder onderzoek en behandeladvies door te sturen (kernpunt 2).

Op de polikliniek orthopedie wordt op aandringen van patiënt een MR-arthrografie (magnetic resonance arthrografie) aangevraagd om intra-articulaire schade uit te sluiten. [5] Er worden geen structurele benige of capsulolabrale afwijkingen gevonden. Daarnaast zijn er geen aanwijzingen voor systemische collageenaandoeningen als m. Ehlers-Danlos of m. Marfan. 4-13% van alle mensen heeft gewrichtshypermobiliteit zonder onderliggende systemische ziekte. [4] Wel vallen tijdens lichamelijk onderzoek een positieve sulcustest (2+) en anterior apprehensiontest op (figuur 1 en 2). [5] Hierbij herkent patiënte het instabiele gevoel dat ze ook tijdens het gebruik van de schouder ervaart. Bij de aanvullende Beighton-schaal test scoorde patiënte 5 punten (tabel 1: hyperextensie knie en passieve dorsoflexie vijfde phalanx waren negatief). [4,8] De orthopaed bevestigt daarom diagnose (instabiliteit door hyperlaxiteit) en behandeladvies van de huisarts en verwijst de patiënte naar de fysiotherapeut (kernpunt 3).

Hier worden de dynamische glenohumerale stabilisatoren (scapulothoracaal en rotator-cuff) progressief getraind op kracht, propriocepsis en neuromusculaire controle.

Gedurende het jaar bemerkt patiënte duidelijke verbetering tijdens werk en sport en na afloop van het fysiotherapeutisch programma is zij haast klachtenvrij en uitermate gemotiveerd om het advies van de huisarts t.a.v. het aansluitende onderhoudsprogramma op te volgen.

Figuur 1: Sulcus sign (door inferiore armtractie wordt de sulcus tussen acromion en humeruskop zichtbaar bij hyperlaxiteit)

Figuur 2: Apprehension test (arm in 90º abductie, elleboog in 90º flexie brengen. Via de pols schouder voorzichtig naar 90º exoroteren, hand drukt tegen achterzijde humeruskop in anterieure richting. Bij apprehension: angst voor luxatie, test positief)

Tabel 1: Beighton-schaal

Test	*Score*
Passieve dorsoflexie vijfde phalanx, >90 graden	1 punt per hand
Passief duim tegen ipsilaterale onderarm plaatsen	1 punt per hand (duim raakt onderarm)
Actieve hyperextensie elleboog, >10 graden	1 punt per elleboog
Actieve hyperextensie knie, >10 graden	1 punt per knie
Al staande handen plat op de grond kunnen leggen, met knieën in volledige extensie (flexie romp)	1 punt
Diagnosticeren van hyperlaxiteit bij ≥4 punten	**Totaal maximaal 9 punten**

Patiënt B is een sportieve 17-jarige jongeman die ongelukkig is gevallen tijdens een "skateboard trick". Hij weet niet hoe hij terecht is gekomen, wel dat zijn rechterschouder/arm verschrikkelijk pijn doet. Hij houdt zijn arm in adductie/endorotatiestand en weigert hem te bewegen. Hij is verder gezond en heeft nooit eerder schouderklachten gehad.

Na enig aandringen mag de huisarts de schouder inspecteren. De schouder blijkt geluxeerd (asymmetrie deltoidcontour, prominerend acromion en rechter humeruskop duidelijker palpabel onder het coracoid). [5] Omdat bij anterieure luxatie soms ook nervus axillaris en plexus brachialis letsel optreedt (schattingen variëren van 10-45%) voert de huisarts een beperkt neurologisch sensibiliteitsonderzoek van de laterale deltoideus-regio uit. [3,5] Er worden geen bijzonderheden gevonden. De huisarts overweegt zelf te reponeren, maar besluit (gezien pijn/angst patiënt, zijn relatieve onervarenheid en omdat hij toch röntgenfoto´s wil maken om fracturen uit te sluiten), dat het beter is om de patiënt na uitleg van diagnose en normaal beloop na repositie door te sturen (kernpunt 4).

Op de spoedeisende hulp bevestigen anteroposteriore en scapulolaterale röntgenfoto's de diagnose van de huisarts en worden geen fracturen gevonden (figuur 3 en 4). Na toediening van een intraveneus analgeticum en spierrelaxans wordt de schouder volgens de Hippocratesmethode soepel gereponeerd met continue armtractie. Neurologisch onderzoek laat opnieuw geen bijzonderheden zien. Hierop wordt de patiënt ontslagen en nabehandeld met 2 weken slingimmobilisatie. [3] Bij poliklinische controle, na 6 weken, is er een volledig pijnvrije "range-of-motion" en patiënt wordt niet langer vervolgd.

Nu meldt de patiënt zich bij zijn huisarts met aanhoudende pijnklachten en instabiel gevoel van de rechterschouder bij aantrekken jas en omdoen autogordel. Er zijn geen bewegingsbeperkingen, maar tijdens het lichamelijk onderzoek zijn de klachten gemakkelijk "provoceerbaar" in abductie/exorotatie stand van de schouder (figuur 2). [5] De huisarts bespreekt diagnose (schouderinstabiliteit na traumatische anterieure schouderluxatie), beloop (hoge re-luxatiekans) en behandelmogelijkheden (fysiotherapie?, operatie?) met de patiënt. Ondanks de hoge recidiefkans wordt besloten om eerst d.m.v. fysiotherapie de schouder te stabiliseren. Nog voor aanvang behandeling meldt de patiënt zich met recidiefluxatie op de spoedeisende hulp na onbekende nachtelijke beweging.

Figuur 3, 4: Röntgenfotoreeks anterieure glenohumerale luxatie

De schouder wordt opnieuw gereponeerd (na uitsluiten van fracturen en neurologische schade) en patiënt keert huiswaarts met in endorotatiesling geïmmobiliseerde schouder. Bij poliklinische controle wordt door de orthopeed een MR-arthrografie aangevraagd om intra-articulaire schouderpathologie te beoordelen (kernpunt 5). Hierop worden een anterior-inferiore capsulolabrale avulsie en een klein posterior-lateraal humeruskopdefect geconstateerd. Omdat er gezien jonge leeftijd en sportieve leefstijl een hoge recidiefluxatiekans is (tot 95%), waardoor de kans op begeleidend letsel toeneemt, wordt tot arthroscopische hersteloperatie besloten (kernpunt 6). [2,7] Hierbij wordt het labrum met drie hecht-ankers anatomisch teruggehecht om stabiliteit te herkrijgen.

Postoperatief is de schouder 6 weken nabehandeld met een shoulderimmobilizer. Onder fysiotherapeutische begeleiding is de schoudermobiliteit in 12 weken hersteld met passieve, geleid-actieve, actieve en stabiliserende oefeningen. Waarna kracht- en sport-

specifieke training werden gestart. Uiteindelijk kon pijnvrij uitoefenen van contactsporten na ongeveer 20 weken weer voorzichtig worden opgepakt. Er is geen recidiefluxatie opgetreden. [5]

Een 71-jarige vrouw (**patiënt C**) meldt zich met intense pijn en ondersteunde rechterarm bij haar huisarts. Ze is door een ruk aan de riem van de hond gevallen met uitgestrekte arm. Ze heeft nooit eerder schouderklachten gehad en verkeert in goede lichamelijke conditie, ondanks corticosteroïdgebruik voor longklachten. Mevr. doet aan ochtendgymnastiek en loopt elke dag 3 km met haar hond.

Tijdens inspectie valt meteen een asymmetrische deltoidcontour en prominerende rechterhumeruskop op. Gezien de verdenking op traumatische anterieure rechterschouderluxatie wordt geen bewegingsonderzoek uitgevoerd. Omdat bij ouderen soms begeleidend letsel van de a. axillaris (1-2%) en n. axillaris (10%-45%) optreedt, wordt wel een neurovasculair onderzoek uitgevoerd. [5,7] Hierbij wordt geen sensomotorische uitval (nervus axillaris en plexus brachialis) geconstateerd. Er is geen axillair hematoom, radiale pulsaties zijn palpabel, huidtemperatuur en capillaire refill van arm/hand zijn normaal. [3,9] De huisarts durft de schouder niet zelf te reponeren door de vergrote kans op fracturen (leeftijd en val), bespreekt dit en verwijst de patiënt door naar het ziekenhuis (kernpunt 4).

Hier wordt door de orthopeed met röntgenfoto's de diagnose van de huisarts bevestigd en leeftijdsgebonden begeleidende fracturen (tuberositas major) uitgesloten. [9] Na intraveneuze toediening van analgeticum en spierrelaxans wordt de arm ongecompliceerd gereponeerd met continue armtractie. Na (post-repositie) neurovasculair onderzoek zonder afwijkingen gaat de patiënte met sling huiswaarts. Om schouderstijfheid te voorkomen wordt meteen gestart met progressieve passieve en actieve "range-of-motion" oefeningen onder fysiotherapeutische begeleiding. Na enkele weken vruchteloze therapie vertrouwt de fysiotherapeut de situatie niet en verwijst terug naar de huisarts.

Mevr. klaagt over aanhoudende pijn en onvoldoende kracht om haar rechterarm te heffen. De huisarts constateert inderdaad forse actieve abductie/exorotatiezwakte tijdens het bewegingsonderzoek en is bang dat ze tijdens haar val ook schouderspier(en) heeft gescheurd. Het risico op rotator-cuffletsel na luxatie is hoog bij patiënten >40 jaar (35-86%) en gezien het corticosteroïdgebruik van mevr. een zeer waarschijnlijke diagnose. [9,10] Omdat beloop en behandelmogelijkheden (fysiotherapie en pijnstilling?, operatieve cuffhechting?) zonder aanvullende diagnostiek moeilijk zijn aan te geven, besluit de huisarts opnieuw door te verwijzen.

De door de orthopeed aangevraagd echo (kernpunt 5) laat een onherstelbare massale cuffruptuur zien. Verschillende operatieve mogelijkheden (partieel cuffherstel, debridement, peestransposities en hemiprothese) worden overwogen. Gezien leeftijd, actieve levensstijl, cuffdeficiëntie met intacte deltoideus, wordt gekozen voor een omgekeerde totaleschouderartroplastiek. Hiermee wordt de meeste pijnreductie en schouderelevatieherstel verwacht (kernpunt 7). [11,12]

Het postoperatieve herstel verliep complicatieloos en patiënte oefende 20 weken "range-of-motion", kracht en functie onder fysiotherapeutische begeleiding. Actieve schouderelevatie verbeterde tot 138 en exorotatie tot 40 graden met minimale pijn.

Beschouwing

De meest voorkomende schouderluxatie is de anterieure luxatie (85-98%). Hierbij kan schade van labrum, rotator-cuff, benige-, caspsuloligamentaire- en neurovasculaire structuren ontstaan. Voordat tot repositie van de schouderluxatie wordt overgegaan dient anterieure luxatie te worden bevestigd en fracturen, neurologische en afhankelijk van leeftijd ook vasculaire schade uitgesloten te worden m.b.v. anamnese, lichamelijk onderzoek en een röntgenreeks. Indien de foto's geen fracturen aantonen en repositie ongecompliceerd verloopt is geen controleröntgenreeks nodig en wordt het vervolgtraject opgestart.

De vervolgbehandelingen zijn divers en afhankelijk van luxatie-oorzaak, leeftijd, activiteitenniveau en ernst van schade door luxatie. Indien operatie wordt overwogen kan begeleidende intra-articulaire schade aangetoond worden door MR-arthrografie. Bij verdenking op periarticulair rotator-cuffletsel (incidentie 35-86% bij patiënten >40 jaar) wordt een echo aangevraagd.

Het doel van conservatieve of operatieve behandeling is een zo pijnloos, functioneel onbeperkt en stabiel mogelijk schoudergewricht te bewerkstelligen. Soms is conservatieve nabehandeling met kortdurende immobilisatie in endorotatiesling (eventueel met fysiotherapie om schouderstijfheid te voorkomen) voldoende, maar bij instabiliteit door schouderhyperlaxiteit wordt een goed resultaat behaald (83-88%) met een uitgebreid fysiotherapeutisch stabiliserend programma van 1 jaar en aanvullend onderhoudsprogramma.

Meestal is echter operatief ingrijpen noodzakelijk om letsel aan statische en/of dynamische stabiliseerders te herstellen en reluxatie te voorkomen. Terwijl bij jonge actieve patiënten al bij een eerste recidiefluxatie (recidiefluxatiekans tot 95%) een stabiliserende operatie wordt gepland is bij uitgebreidere pathologie (zoals massaal cuffletsel) soms zelfs een omgekeerde totaleschouderartroplastiek noodzakelijk.

Conclusie

Zoals uit bovenstaande casussen blijkt is doorverwijzing naar een schouderspecialist alleen geïndiceerd bij patiënten met traumatische anterieure schouderluxatie. Enerzijds betreft dit primaire luxaties waarbij een juiste follow-up met röntgenfoto's in een ziekenhuis onontbeerlijk is. Anderzijds betreft het problematiek rondom beperkt herstel van "range-of-motion", totale afwezigheid van kracht of aanhoudende schouderinstabiliteit, waarbij snelle doorverwijzing t.b.v. aanvullende diagnostiek en behandeling verdere schade voorkomt. Instabiliteit door hyperlaxiteit kan door de huisarts worden vastgesteld

m.b.v. de sulcustest, anterior apprehensiontest en Beighton-schaal. Doorverwijzen is hier alleen nuttig indien uitgebreide fysiotherapie niet helpt.

Tabel 2: De kern

1)	Bij (anterieure) instabiliteit door schouderhyperlaxiteit wordt een conservatief stabiliserend fysiotherapeutisch programma van 1 jaar met onderhoudsbehandeling geadviseerd.
2)	Bij schouderhypermobiliteit is rotator-cuff letsel of intra-articulaire schade niet waarschijnlijk en als er al capsulolabrale afwijkingen zijn dan zijn de operatieve resultaten zeer matig. Doorverwijzen naar een orthopaedisch chirurg met specifieke kennis van schouderpathologie is alleen nuttig indien fysiotherapie (kernpunt 1) niet helpt.
3)	Om de diagnose instabiliteit door schouderhyperlaxiteit te bevestigen zou de huisarts tijdens het lichamelijk onderzoek de sulcustest, anterior apprehensiontest en Beighton-schaal uit kunnen voeren.
4)	Bij anterieure luxatie dienen fracturen, neurologische en afhankelijk van leeftijd ook vasculaire schade uitgesloten te worden. Doorverwijzing naar een schouderspecialist is hier altijd geïndiceerd.
5)	Intra-articulaire schouderpathologie door anterieure luxatie wordt aangetoond met een MR-arthrografie maar bij verdenking van periarticulair letsel (rotator-cuff) wordt een echo verricht.
6)	Bij jonge actieve patiënten wordt gezien hun hoge recidiefkans (tot 95%) na traumatische anterieure schouderluxatie bij voorkeur snel tot een scopische anatomische hersteloperatie overgegaan. Snelle doorverwijzing naar de tweede lijn bij aanhoudende instabiliteit voorkomt verdere schade door reluxaties.
7)	Bij oudere patiënten kan bij beperkte vordering van de revalidatie, met name in het herstel van “range-of-motion” en kracht, een prompte doorverwijzing voor nadere diagnostiek erg zinvol zijn gezien de operatiemogelijkheden variërend van cuffhechting tot omgekeerde totaleschouderartroplastiek.

Referenties

1. Gaskill TR, Taylor DC, Millett PJ. Management of multidirectional instability of the shoulder. J Am Acad Orthop Surg. 2011;19:758-767.
2. Sileo MJ, Joseph S, Nelson CO, Botts JD, Penna J. Management of acute glenohumeral dislocations. Am J Orthop. 2009;38:282-290.
3. Richtlijn NOV: Acute primaire schouderluxatie. Diagnostiek en behandeling. Nijmegen: Nederlandse Orthopaedische Vereniging; 2005.
4. Johnson SM, Robinson CM. Shoulder Instability in patients with joint hyperlaxity. J Bone Joint Surg Am. 2010;92:1545-57.
5. Wang RY, Arciero RA, Mazzocca AD. The recognition and treatment of first-time shoulder dislocation in active individuals. J Orthop Sports Phys Ther. 2009;39:118-123.
6. Winters JC, van der Windt DA, Spinnewijn WE, de Jongh AC, van der Heijden GJ, Buis PA, et al. NHG-standaard schouderklachten (tweede herziening). Huisarts Wet 2008:51(11):555-565.
7. Cutts S, Prempeh M, Drew S. Anterior shoulder dislocation. Ann R Coll Surg Engl. 2009;91:2-7.
8. Wolf JM, Cameron KL, Owens BD. Impact of joint laxity and hypermobility on the musculoskeletal system. J Am Acad Orthop Surg. 2011;19:463-471.
9. Murthi AM, Ramirez MA. Shoulder dislocation in the older patient. J Am Acad Orthop Surg. 2012;20:615-622.
10. Dodson CC, Cordasco FA. Anterior glenohumeral joint dislocations. Orthop Clin N Am. 2008;39:507-518.
11. Ramirez MA, Ramirez J, Murthi AM. Reverse total shoulder arthroplasty for irreparable rotator cuff tears and cuff tear arthropathy. Clin Sports Med. 2012;31:749-759.
12. Delaney RA, Lin A, Warner JJP. Nonarthroplasty options for the management of massive and irreparable rotator cuff tears. Clin Sports Med. 2012;31:727-748.

Chapter | 3

MR arthrography of traumatic anterior shoulder lesions showed modest reproducibility and accuracy when evaluated under clinical circumstances

Susan van Grinsven, Oscar Kesselring, Hester van Wassenaer-van Hall,
Robert Lindeboom, Cees Lucas, Corné van Loon

Archives of Orthopaedic and Trauma Surgery 2007;127:11-17

Abstract

Introduction: We examined the reproducibility and accuracy of high-field MRA in traumatic anterior shoulder instability under conditions resembling clinical practice and assessed the influence of observer experience.

Materials and methods: Two radiologists with different experience levels evaluated 61 MRA's. Assessment was independent, blinded and non-sequential. For 40 MRA's, surgical reports were available to assess diagnosis accuracy and influence of observer experience. The assessed lesions were cuff lesions, Hill–Sachs lesions, bony and classic Bankart lesions, greater humeral tuberosity fractures, SLAP lesions and joint capsule lesions. Reproducibility was quantified using kappa coefficients. Accuracy was evaluated with sensitivity and specificity rates, positive and negative predictive values. Differences in the percentage of correctly diagnosed MRA's between the radiologists were tested using the McNemar test for paired proportions.

Results: Inter-observer k-values ranged from 0.03 for joint capsule lesions to 0.45 for humeral head lesions. The overall kappa was 0.21 (95% CI; 0.12–0.30). We also observed markedly lower sensitivity and specificity rates than those reported in the literature for most lesions. The more experienced radiologist correctly diagnosed 78.9% of all lesions compared to 65.4% for the less experienced radiologist ($p < 0.001$; McNemar test).

Conclusion: MRA-interpretations of traumatic anterior shoulder instability should be regarded with caution in clinical practice. The experience level of radiologists can affect reproducibility and accuracy.

Introduction

The shoulder joint is the most mobile joint in the human body. Unfortunately, this mobility threatens the joint's stability. Most dislocations of the shoulder joint are in anterior direction (85%) and of traumatic origin (96%). Anterior lesions such as Bankart, joint capsule or Hill–Sachs lesions are usually seen in younger patients, while rotator cuff tears or greater humeral tuberosity fractures are more often observed in older patients. [20]

In subtle forms of instability, diagnoses or instability directions are not clearly assessed with patient's history, physical examination and standard X-ray evaluation. Therefore, additional imaging techniques were recommended. [6] Magnetic resonance arthrography (MRA) is the technique of choice for detecting these subtle lesions [2-5, 7,8,10-19,21-26,29-31] and is therefore used to confirm diagnosis and plan surgical stabilization.

Although adequate accuracy rates of the MRA were observed in subjects with traumatic shoulder instability, evaluation was mostly done under ideal circumstances, i.e. assessment by an experienced observer or in consensus between two observers. [2-5,7,8,11,12,14-16,19,21-26,29-31] In normal clinical practice, however, radiologists with different experience levels assess MRA's independently and often have to deal with time constraint. Reproducibility and accuracy of MRA assessments in research settings may therefore be overestimated.

In view of the usefulness of MRA to select subjects eligible for stabilizing surgery, we examined the reproducibility and accuracy of high-field MRA in traumatic anterior shoulder instability under conditions resembling clinical practice. We also assessed the influence of observer experience on MRA assessment accuracy.

Materials and methods

All patients who were referred after standard shoulder X-ray (AP-internal, external and transscapular view) [28] and underwent an 1.5 T MRA in our secondary care setting from January 2001 to June 2004 were considered for enrolment in this diagnostic retrospective cohort study. Those included were skeletally mature patients with traumatic anterior shoulder instability. Patients were excluded if they had nontraumatic / traumatic inferior or posterior shoulder instability, refused surgery or had previous shoulder surgery.

Arthrography was performed less than 30 min before the MR imaging to obtain optimum imaging. After an anaesthetising injection (Lidocaine 1%, B. Braun Melsungen AG, Melsungen, Germany) a 21 gauge needle was fluoroscopically guided into the superior medial quadrant of the humeral head. Confirmation of correct intra-articular needle placement was obtained by injecting 2–3 cc iodinated contrast agent (Xenitix 300, Guerbet Nederland B.V., Gorinchem, The Netherlands), after which 14–16 cc diluted gadolinium complex (Dotarem, Guerbet Nederland B.V., Gorinchem, The Netherlands) was instilled

in the glenohumeral joint. MR imaging was performed (Table 1) with an 1.5 T ACS Intera Gyroscan system and a Synergy flex-M surface shoulder coil (Philips, Best, The Netherlands). We did not exclude sub-optimal MRA's, due to time constraint. Time pressure caused by the high amount of daily MR-images in our hospital, sometimes leads to sub-optimal number of slices and skipping sequences. In fact, the ABER position view was used in none of the cases. In this abducted externally rotated arm position, visualization of the anterior labral–ligamentous complex and lesion detection rate are improved by traction. [10,28]

To assess reproducibility and accuracy, two radiologists, with different experience levels, assessed all included MRA's independently and non-sequentially, using an eight-item scoring list (Table 2). Observer 1 was a CT/MR radiologist with 17 years of experience, of which 6 years with MRA. Observer 2 was a general radiologist with 10 years of experience.

Table 1: MR imaging (1.5 Tesla) protocol

Sequence	***Gradient echo / T1-weighted (T1W/FFE/3D)***	***Turbo spin echo / T2-weighted (T2W/TSE)***	***Proton density weighted / Spin echo (PDW/SE)***	***Turbo spin echo / T1-weighted (T1W/TSE)***	***Turbo spin echo / T1-weighted (T1W/TSE)***
Orientation	Oblique coronal	Oblique coronal	Sagittal	Axial	ABER position[a]
Fat suppressed	Yes	Yes	No	Yes	Yes
Time to repeat (ms)	23	3269	1800	475	475
Time to echo (ms)	9.5	70	25	18	18
Flip angle (°)	20	90	90	90	90
Slice thickness (mm)	4	3.5	3.5	3.5	3.5
Slice gap (mm)	--	0.35	0.35	0.35	0.35
NEX[b]	3	4	2	3	3
Duration (min)	4.10	3.25	4.51	5.28	5.28
FOV[c] (mm)	170	180	130	180	180
Matrix size	304 x 228	256 x 205	288 x 202	304 x 212	256 x 205
Reconstructionmatrix	512	512	512	512	512

[a]*: the optional abduction external rotation view improves visualization of the anterior labroligamentous complex*

[b]*: number of excitations*

[c]*: field of view*

To assess the accuracy of MRA assessments, arthroscopy and/or open surgery reports served as the reference standard and were interpreted in consensus by two reviewers (SvG, CvL) using the same scoring list. All raters were blinded to the patients' data, clinical history (except for age, gender and inclusion criteria), MRA findings and/or surgical reports.

The pathologies assessed (Table 2) were cuff lesions, Hill–Sachs lesions, bony and classic Bankart lesions, greater humeral tuberosity fractures, SLAP lesions and joint capsule lesions. [6,20,27,28] The spectrum of pathology of the MRA's ranged from no abnormalities to the presence of multiple lesions.

Statistical analysis

Inter-observer agreement, regarding the presence of a given lesion on the MRA's, was expressed using kappa coefficients [9] and absolute percentages of agreement. Kappa coefficients can be categorised as follows: 0.00-0.20, poor; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial and 0.81-1.00, almost perfect. In addition, a kappa with its 95% confidence interval was calculated, expressing the overall reproducibility of the MRA assessments (427 pairs of observations: MRA's x seven lesion types).

The accuracy of the MRA assessments, as confirmed by the surgical report, was expressed in sensitivity and specificity rates and positive and negative predicted values. The influence of experience of the two radiologists on diagnoses accuracy was evaluated by comparing the percentages of correct diagnosis per lesion type and over all lesions types in surgically confirmed cases. Differences between the radiologists in the percentages of correctly diagnosed lesions on the MRA's were tested using the McNemar test for paired proportions. [1] Differences were considered significant at the 5% level. Statistical evaluations were carried out using SPSS 11.5 software. Post hoc sample size calculations indicated that with 280 pairs of observations in our sample we could statistically detect, with a power of 80% and a type I error rate of 5%, a 15% difference in correctly diagnosed lesions between the radiologists when the proportion of discordant pairs is conservatively assumed 0.65 and the method of analysis is McNemar test of equality of paired proportions.

Table 2: Frequency and reproducibility of the MRA pathologies (N = 61)

Items standard scoring list	*Frequency R1/R2*	*Kappa*	*Absolute percentage of agreement (%)*
Cuff lesion:		0.19	64
Not present	53/35		
Partial tear	07/24		
Full thickness tear	01/02		
Humeral head lesion:		0.45	74
Not present	38/15		
Degeneration	00/21		
Hill-Sachs	23/25		
Anterior inferior glenoid lesion:		0.11	73
Not present	54/40		
Degeneration	03/05		
Bony Bankart	04/16		
Greater humeral tuberosity fractures:[a]		---	97
Not present	61/59		
Present	00/02		
Anterior inferior labrum lesion:		0.15	57
Not present	24/15		
Degeneration	06/17		
Classic Bankart	31/29		
SLAP lesion:		0.04	57
Not present	53/35		
Present	08/26		
Joint capsule lesion:		0.03	51
Not present	60/30		
Capsule redundancy	01/13		
Anterior stripping / Tear	00/18		
Quality of MRA:			
Good	44/52		
Moderate	15/09		
Poor	02/00		
Overall kappa (N=427 pairs)		**0.21**	
95% confidence interval overall kappa		**[0.12–0.30]**	
Overall absolute % agreement			**67**

SLAP: superior labrum anterior-to-posterior lesion, R1: radiologist 1, R2: radiologist 2, [a]: calculation of kappa not possible due to empty cells in 2x2 table

Results

Only 61 MRA's (40 men and 21 women), of the 103 MRA's considered for enrolment in this study, met the inclusion criteria. The mean age at MRA-time was 29 years (range 16-53). Forty patients (27 men and 13 women) underwent surgical stabilization. Arthroscopy was performed in 27 cases and open surgery in 20 cases. The median time from MRA to surgery was 68 days (range 8-513).

Table 2 summarizes the frequencies and reproducibility for each lesion as calculated, using the MRA ratings of both radiologists. The kappa values ranged from 0.03 for joint capsule lesions to 0.45 for Hill-Sachs lesions. The overall kappa, calculated using 427 pair-wise ratings, was 0.21 (95% CI; 0.12-0.30).

Table 3 summarizes the frequency of each lesion and the accuracy of the MRA's in 40 surgery confirmed cases. For the computation of accuracy rates, we considered structures not mentioned in the surgical reports to be normal. The sensitivity rates for radiologist 1 (range 7-100%) were sometimes lower than those of radiologist 2 (range 29-100%). Radiologist 1, however, had higher specificity rates and generally higher positive and negative predicted values than the less experienced radiologist. The prevalence of cuff lesions and greater humeral tuberosity fractures in the surgical reports was, however, rather low. The overall sensitivity, regardless of lesion type, for radiologist 1 was 50.0% (33 of 66 lesions present) with a specificity of 87.9%. The reviewers rated the majority of the surgical reports, based on completeness and detailed description, as moderate.

The percentages of correctly diagnosed lesions, in the 40 cases with surgical confirmation, are displayed in Table 4. The more experienced radiologist 1 had higher percentages (range 65.0-100%) of correctly diagnosed lesions (true positive and true negative) for all lesion types, than radiologist 2 (range 47.5-95.0%). For three lesion types this difference was significant ($p < 0.05$; McNemar test). The overall percentage of correctly diagnosed lesions on the MRA's, based on 280 pair-wise compared diagnoses scores, was 78.9% for radiologist 1 and 65.4% for radiologist 2 ($p < 0.001$; McNemar test).

Table 3: Accuracy of MRA diagnoses in surgery confirmed cases (N = 40)

Items standard scoring list		*Frequency R1/R2/ Surg*	*Sensitivity R1/R2*	*Specificity R1/R2*	*Positive predicted value R1/R2*	*Negative predicted value R1/R2*
Cuff lesion:			100/100	82/62	13/06	100/100
	Not present	32/24/08				
	Partial tear	07/14/00				
	Full thickness tear	01/02/01				
Humeral head lesion:			80/60	73/70	50/40	92/84
	Not present	24/09/10				
	Degeneration	00/16/01				
	Hill-Sachs	16/15/10				
Anterior inferior glenoid lesion:			25/50	94/69	33/15	92/93
	Not present	34/22/09				
	Degeneration	03/05/06				
	Bony Bankart	03/13/04				
Greater humeral tuberosity fractures:			NA	NA	NA	NA
	Not present	40/38/40				
	Present	00/02/00				
Anterior inferior labrum lesion:			69/62	64/55	83/78	44/35
	Not present	14/08/03				
	Degeneration	02/09/07				
	Classic Bankart	24/23/29				
SLAP lesion:			29/29	85/52	29/11	85/77
	Not present	33/22/13				
	Present	07/18/07				
Joint capsule lesion:			07/73	100/48	100/46	64/75
	Not present	39/16/07				
	Capsule redundancy	01/09/11				
	Anterior stripping / Tear	00/15/04				

Quality of surgical report:			
	Good	--/--/06	
	Moderate	--/--/29	
	Poor	--/--/05	
Overall sensitivity		**50.0/60.6**	
Overall specificity			**87.9/66.8**

NA: not applicable since fractures of the humeral head were not observed, R1: radiologist 1, R2: radiologist 2, Surg: surgical reports (arthrotomy or arthroscopy), only structures that were mentioned in the surgical reports are displayed

Table 4: Percentage correctly diagnosed per lesion type for radiologist 1 and radiologist 2 as confirmed by the surgical reports (N = 40)

	Radiologist 1 (%)	*Radiologist 2 (%)*	*P*[a]
Cuff lesion	33/40 (82.5)	25/40 (62.5)	0.04
Humeral head lesion	30/40 (75.0)	27/40 (67.5)	0.51
Anterior inferior glenoid lesion	35/40 (87.5)	27/40 (67.5)	0.04
Greater humeral tuberosity fractures	40/40 (100)	38/40 (95.0)	NA
Anterior inferior labrum lesion	27/40 (67.5)	24/40 (60.0)	0.63
SLAP lesion	30/40 (75.0)	19/40 (47.5)	0.01
Joint capsule lesion	26/40 (65.0)	23/40 (57.5)	0.68
Overall percentage correct (N=280 pairs)	**(78.9)**	**(65.4)**	**0.00**

NA: not applicable, McNemar test p-value could not be determined due to empty cells in 2x2 table, %: percentage, [a]: McNemar test for equality of paired proportions

Discussion

We examined the reproducibility and accuracy of MRA in patients with anterior shoulder instability under a research regimen resembling clinical practice.

Although kappa values depend on prevalence, we observed markedly lower k-values for five of the six lesions (range 0.03-0.19) than reported in former research (range 0.20-1.00). [4,10,13,15,17,18] The overall kappa (agreement regardless of lesion type) was only in the "fair" range (0.21).

Pair-wise comparison of all MRA's revealed that the more experienced radiologist 1 significantly diagnosed 14% more lesions correctly than radiologist 2 (78.9 versus 65.4%; $p < 0.001$).

With regard to MRA accuracy, we only discuss the results of the more experienced radiologist and exclude cuff and humeral head lesions from this discussion because of low prevalence in the surgical reports. Partial (intrasubstance) tears are, unless the tendon is incised, difficult to confirm with open surgery or arthroscopy. [32]

The sensitivity and specificity for Hill-Sachs lesions in our study were 80/73%. The results were supported by Bitzer et al., but other authors reported percentages of 100%. [18,25] The low sensitivity of 25% for bony Bankart lesions was supported by Sano et al. Other researchers found substantially higher percentages (range 67-100%). [5, 18, 25]

The sensitivity and specificity for classic Bankart lesions were 69/64%. Others consistently reported percentages well over 80%. [2,5,7,8,11,13,18,23-25,29,30] In line with our study, Cvitanic et al. found a low sensitivity result of 48% when not using the ABER position sequence. We found a sensitivity and specificity for SLAP lesions of 29/85%. Although the specificity was consistent with earlier reports (range 79-98%), the

sensitivity was much lower (range 50-92%). [3,4,17,31] The low sensitivity for joint capsule lesions of 7% was not supported by other reports (range 75-100%). [5,11,18]

In spite of the significantly higher percentage of correctly diagnosed lesions by radiologist 1 we generally observed lower reproducibility and accuracy results in our study, resembling clinical practice, than reported in former reports. [2-5,7,8,10,11,13,15, 17,18,23-25,29-31]

Study limitations (retrospective character, moderate surgical report quality, structures not mentioned were considered normal and different surgery techniques) may have resulted in biased estimates. However, arthroscopy and open surgery enables the surgeon to fully examine the anterior part of the joint. It is therefore unlikely that anterior joint capsule lesions, SLAP lesions or bony and classic Bankart lesions were missed and if present, they would be mentioned in the surgical reports.

Being a teaching hospital with 750 beds and an adherence of 325.000 inhabitants, these results forced us to recommend changes in our MRA protocol. We recommended standard feedback of the surgical findings. To confirm diagnosis and to plan stabilizing surgery, assessment of MRA's should only be performed in consensus by experienced raters. We further recommended standard use of the ABER position view and optimal amount of slices. The absence of the ABER position view in our study could have been an explanation for the lower reproducibility and accuracy results of some of the lesions observed. [10,28] The exact influence, however, of time constraint and independent assessment on reproducibility and accuracy is still subject for further prospective diagnostic research.

Conclusion

Although MRA is the technique of choice for detecting subtle lesions associated with shoulder instability, MRA-interpretations in clinical practice should be regarded with caution by orthopaedic surgeons. The experience level of radiologists can affect reproducibility and accuracy.

References

1. Altman DG. Statistics in medical journals: developments in the 1980s. Stat Med 1991;10(12):1897-1913.
2. Bachmann G, Bauer T, Jürgensen I, Schwab J, Weimar B, Rau WS. Diagnostische Sicherheit und therapeutische Relevanz von CT-Arthrographie und MR-Arthrographie der Schulter. Fortschr Röntgenstr 1998;168(2):149-156.
3. Bencardino JT, Beltran J, Rosenberg ZS, Rokito A, Schmahmann S, Mota J, et al. Superior labrum anterior-posterior lesions: diagnosis with MR arthrography of the shoulder. Radiology 2000;214:267-271.
4. Binkert CA, Zanetti M, Gerber C, Hodler J. MR arthrography of the glenohumeral joint: two concentrations of Gadoteridol versus Ringer solution as the intraarticular contrast material. Radiology 2001;220:219-224.
5. Bitzer M, Nasko M, Krackhardt T, Schick F, Schöber W, Wiskirchen J, et al. Direkte CT-Arthrographie versus direkte MR-Arthrographie bei chronischer Schulterinstabilität: Ein Methodenvergleich nach Einführung der Multidetektor-CT-Technik. Fortschr Röntgenstr 2004;176:1770-1775.
6. Blum A, Coudane H, Molé D. Gleno-humeral instabilities. Eur Radiol 2000;10:63-82.
7. Chandnani VP, Yeager TD, DeBerardino T, Christensen K, Gagliardi JA, Heitz DR, et al. Glenoid labral tears: prospective evaluation with MR imaging, MR arthrography, and CT arthrography. AJR Am J Roentgenol 1993;161:1229-1235.
8. Chandnani VP, Gagliardi JA, Murnane TG, Bradley YC, DeBerardino TA, Spaeth J, et al. Glenohumeral ligaments and shoulder capsular mechanism: evaluation with MR arthrography. Radiology 1995;196:27-32.
9. Cohen J. A coeffcient of agreement for nominal scales. Educ Psychol Meas 1960;20:37-46.
10. Cvitanic O, Tirman PF, Feller JF, Bost FW, Minter J, Carroll KW. Using abduction and external rotation of the shoulder to increase the sensitivity of MR arthrography in revealing tears of the anterior glenoid labrum. AJR Am J Roentgenol 1997;169:837-844.
11. Drescher R, von Rothenburg T, Ludwig J, Köster O, Schmid G. Direkte MR-Arthrographie des schultergelenks mit maximaler Kapseldistension zur OP-Planung. Fortschr Röntgenstr 2004;176:1469-1474.
12. Ferrari FS, Governi S, Burresi F, Vigni F, Stefani P. Supraspinatus tendon tears: comparison of US and MR arthrography with surgical correlation. Eur Radiol 2002;12:1211-1217.
13. Flannigan B, Kursunoglu-Brahme S, Snyder S, Karzel R, Del Pizzo W, Resnick D. MR arthrography of the shoulder: comparison with conventional MR imaging. AJR Am J Roentgenol 1990;155:829-832.
14. Funke M, Kopka L, Vosshenrich R, Oestmann W, Grabbe E. MR arthrography in the diagnosis of rotator cuff tears. Standard spin-echo alone or with fat suppression. Acta Radiologica 1996; 37:627-632.
15. Guntern DV, Pfirrmann CW, Schmid MR, Zanetti M, Binkert CA, Schneeberger AG, et al. Articular cartilage lesions of the glenohumeral joint: diagnostic effectiveness of MR arthrography and prevalence in patients with subacromial impingement syndrome. Radiology 2003;226:165-170.

16. Hodler J, Kursunoglu-Brahme S, Snyder SJ, Cervilla V, Karzel R, Schweitzer ME, et al. Rotator cuff disease: assessment with MR arthrography versus standard MR imaging in 36 patients with arthroscopic confirmation. Radiology 1992;182:431-436.
17. Jee W, McCauley TR, Katz LD, Matheny JM, Ruwe PA, Daigneault JP (2001) Superior labral anterior posterior (SLAP) lesions of the glenoid labrum: reliability and accuracy of MR arthrography for diagnosis. Radiology 2001;218:127-132.
18. Loew R, Kreitner KF, Runkel M, Zoellner J, Thelen M. MR arthrography of the shoulder: comparison of low-field (0.2 T) vs high-field (1.5 T) imaging. Eur Radiol 2000;10:989-996.
19. Magee T, Williams D, Mani N. Shoulder MR arthrography: which patient group benefits most. AJR Am J Roentgenol 2004;183:969-974.
20. Matsen FA, Rockwood CA. Glenohumeral instability. The Shoulder. Matsen FA, Rockwood CA, editors. WB Saunder Company, Philadelphia 1998;611-745.
21. Meister K, Thesing J, Montgomery WJ, Indelicato PA, Walczak S, Fontenot W. MR arthrography of partial thickness tears of the undersurface of the rotator cuff: an arthroscopic correlation. Skeletal Radiol 2004;33:136-141.
22. Palmer WE, Brown JH, Rosenthal DI. Rotator cuff: evaluation with fat-suppressed MR arthrography. Radiology 1993;188:683-687.
23. Palmer WE, Brown JH, Rosenthal DI. Labral–ligamentous complex of the shoulder: evaluation with MR arthrography. Radiology 1994;190:645-651.
24. Palmer WE, Caslowitz PL. Anterior shoulder instability: diagnostic criteria determined from prospective analysis of 121 MR arthrograms. Radiology 1995;197:819-825.
25. Parmar H, Jhankaria B, Maheshwari M, Singrakhia M, Shanbag S, Chawia A, et al. Magnetic resonance arthrography in recurrent anterior shoulder instability as compared to arthroscopy: a prospective comparative study. J Postgrad Med 2002;48:270-273.
26. Roger B, Skaf A, Hooper AW, Lektrakul N, Yeh L, Resnick D. Imaging findings in the dominant shoulder of throwing athletes: comparison of radiography, arthrography, CT arthrography, and MR arthrography with arthroscopic correlation. AJR Am J Roentgenol 1999;172:1371-1380.
27. Sailer J, Imhof H. Schultergelenkinstabilität. Radiologe 2004;44:578-590.
28. Sanders TG, Morisson WB, Miller MD. Imaging techniques for the evaluation of glenohumeral instability. Am J Sports Med 2000;28:414-434.
29. Sano H, Kato Y, Haga K, Itoi E, Tabata S. Magnetic resonance arthrography in the assessment of anterior instability of the shoulder: comparison with double-contrast computed tomography arthrography. J Shoulder Elbow Surg 1996;5:280-285.
30. Sasaki T, Saito Y, Yodono H, Prado GL, Miura H, Itabashi Y, et al. Technical note. Labral–ligamentous complex of the shoulder. Evaluation with double oblique axial MR arthrography. Acta Radiol 2003;44:435-439.
31. Waldt S, Burkart A, Lange P, ImhoV AB, Rummeny EJ, Woertler K. Diagnostic performance of MR arthrography in the assessment of superior labral anteroposterior lesions of the shoulder. AJR Am J Roentgenol 2004;182:1271-1278.
32. Zlatkin MB. Rotator cuff disease. MRI of the shoulder. Lippincott Williams & Wilkins, Wolters Kluwer Company, Philadelphia, 2003;117–175.

Chapter | 4

Does the experience level of the radiologist, assessment in consensus, or the addition of the abduction and external rotation view improve the diagnostic reproducibility and accuracy of MRA of the shoulder?

Susan van Grinsven, Femke Hagenmaier, Corné van Loon,
Maarten van Gorp, Marjolijn van Kints, Albert van Kampen

Clinical Radiology 2014;69:1157-1164

Abstract

Aim: To prospectively evaluate the influence of observer experience, consensus assessment, and abduction and external rotation (ABER) view on the diagnostic performance of magnetic resonance arthrography (MRA) in patients with traumatic anterior shoulder instability (TASI).

Materials and methods: Fifty-eight MRA examinations (of which 51 had additional ABER views) were assessed by six radiologists (R1-R6) and three teams (T1-T3) with different experience levels, using a seven-lesion standardized scoring form. Forty-five out of 58 MRA examination findings were surgically confirmed. Kappa coefficients, sensitivity, specificity, and differences in percent agreement or correct diagnosis (p-value, McNemar test) were calculated per lesion and overall per seven lesion types to assess diagnostic reproducibility and accuracy.

Results: Overall kappa ranged from poor ($k = 0.17$) to moderate ($k = 0.53$), sensitivity from 30.6-63.5%, and specificity from 73.6-89.9%. Overall, the most experienced radiologists (R1-R2) and teams (T1-T2) agreed significantly more than the lesser experienced radiologists (R3-R4: $p = 0.014$, R5-R6: $p = 0.018$) and teams (T2-T3: $p = 0.007$). The most experienced radiologists (R1, R2, R3) and teams (T1, T2) were also consistently more accurate than the lesser experienced radiologists (R4, R5, R6) and team (T3). Significant differences were found between R1-R4 ($p = 0.012$), R3-R4 ($p = 0.03$), and T2-T3 ($p = 0.014$). The overall performance of consensus assessment was systematically higher than individual assessment. Significant differences were established between T1-T2 and radiologists R3-R4 ($p = 0.001$, $p < 0.001$) and between T2 and R3 ($p = 0.050$) or R4 ($p < 0.001$, $p = 0.001$). No overall significant differences were found between the radiologists' assessments with and without ABER.

Conclusion: The addition of ABER does not significantly improve overall diagnostic performance. The radiologist's experience level and consensus assessment do contribute to higher reproducibility and accuracy.

Introduction

Traumatic anterior dislocation is the most frequent cause (95%) of shoulder instability, and incidences vary between 0.08-0.24 per 1000 person-years. [1-5] Well-known associated rotator cuff, osseous, cartilage, and labroligamentous complex lesions [1,3,4,6-12] are often subtle and difficult to diagnose based on patient history and physical examination alone. Therefore, diagnosis and appropriate surgery planning is frequently dependent on additional imaging. [3,11-14]

Currently, magnetic resonance arthrography (MRA) is considered the "pre-arthroscopy" diagnostic reference standard. [1-3,7,10-12,14-27] Depending on the examined lesion type, [12,19,20,28-35] MRA sensitivities and specificities range from 79-100% and 85-100%, [36] respectively, although much lower values are reported as well. [3,6,7,15,37] The diagnostic MRA reproducibility rates tend to be lower than the accuracy rates and are highly variable. Kappa's ranging from "poor" (-0.03) to "almost perfect agreement" (0.84) are described for instability related lesions. [3,6,19,25,26,31,34,37-40]

Addition of the abduction and external rotation (ABER) view is presumed to optimize the diagnostic performance of MRA by improving the visualization of rotator cuff partial thickness tears and any present lesion in the anterior labroligamentous complex through distraction and increased penetration of contrast material into a tear. [2,6,12,14,18,24,27,30-32,34,37] However, its possible beneficial effect on diagnostic reproducibility and accuracy is disputed in the present literature, as both significant superior, [2,14,23,34,38] significant inferior, [27,34] and no significant difference in reproducibility and accuracy [25,26,33,34] have been described.

As a possible explanation for these conflicting reproducibility and accuracy rates, it has been mentioned that the diagnostic performance of shoulder MRA, might depend on the experience level of the interpreting radiologists. [3,4,7,31,36,41] The literature also suggests that individual instead of consensus assessment could be of influence too, as radiologists can learn from one another by discussing shoulder lesions that are difficult to diagnose at MRA. [3,19] However, there are no sound prospective comparative diagnostic studies on these topics to confirm this. As systematic accurate MRA diagnosis of each separate radiologist is only achievable through high reproducibility, the aim of the present study was to evaluate the influence of observer experience, assessment in consensus, and the additional value of ABER on the diagnostic reproducibility and accuracy of high-field MRA in patients with traumatic anterior shoulder instability (TASI). The hypothesis was that experience level, consensus assessment, and ABER does increase diagnostic performance.

Materials and methods

All patients who visited Rijnstate Hospital's outpatient clinic between 2007 and 2011 with suspected TASI, after reduction of traumatic anterior shoulder dislocation, were considered for enrolment in this prospective diagnostic performance study. Rijnstate Hospital is a teaching hospital with 750 beds and an adherence of 325.000 inhabitants. Patients were excluded in case of previous shoulder surgery, general contraindications to MRI and/or contrast agents, skeletal immaturity, and refusal to participate. Ultimately, 58 MRA examinations were included in the study. In 51 of the 58 patients, an ABER view was obtained. Forty-five of the 58 patients underwent surgery, allowing for confirmation of the MRA findings. The 45 surgery confirmed MRA's served as the reference standard. Study protocol implementation and data collection were planned in advance and all members of the orthopaedic and radiological departments were instructed accordingly. The regional ethics committee decided that neither approval nor informed consent were required, however, the study was approved by the local scientific committee of the Rijnstate Hospital.

MRA (index test)

Shoulder MRI was performed according to a well-defined, standardized study protocol (Table 1). For the arthrography, an anterior approach was used to insert a 21 G needle into the inferior or superior-medial quadrant (according to the radiologist's preference) of the humeral head with fluoroscopic guidance. Intra-articular position was confirmed with iodinated contrast agent (2-3 ml iobitridol; 300 mg iodine/ml; Xenitix 300, Guerbet, Gorinchem, The Netherlands) and 14-16 ml diluted gadolinium complex (Artirem, Guerbet, Gorinchem, The Netherlands) was instilled into the glenohumeral joint. To obtain optimum imaging quality, MRI images were acquired within 30 minutes after arthrography. MRI images were produced using one of three MRI machines [1.5 Tesla ACS Intera Gyroscan system/Synergy flex-M surface shoulder coil (2001), which was replaced during the study period (2008) for a 1.5 Tesla Achieva system/Synergy flex-M surface shoulder coil or a 1.5 Tesla Siemens Magnetom Avanto Tim 32 x 8/Small Extremity coil]. Patients were placed in a supine position, the arm slightly abducted and exorotated. For the additional ABER view patients were instructed to place the hand of the affected extremity posterior to the contralateral aspect of the head or neck with the elbow flexed.

Image evaluation and reference standard

The MRA report was send to the orthopaedic surgeon to confirm clinical diagnosis and plan arthroscopic stabilization. All arthroscopies were performed by two orthopaedic surgeons with substantial arthroscopic experience (6 and 8 years) and specialized in

Table 1: MR imaging (1.5 Tesla) protocol

Sequence	***Gradient echo/ T1-weighted (T1W/FFE/3D)***	***Turbo spin echo/ T2-weighted (T2W/TSE)***	***Proton density weighted/ Spin echo (PDW/SE)***	***Turbo spin echo/ T1-weighted (T1W/TSE)***	***Turbo spin echo/ T1-weighted (T1W/TSE)***
Orientation	Oblique coronal	Oblique coronal	Sagittal	Axial	ABER position[a]
Fat suppressed	Yes	Yes	No	Yes	Yes
Time to repeat (ms)	23	3269	1800	475	475
Time to echo (ms)	9.5	70	25	18	18
Flip angle (º)	20	90	90	90	90
Slice thickness (mm)	4	3.5	3.5	3.5	3.5
Slice gap (mm)	---	0.35	0.35	0.35	0.35
NEX[b]	3	4	2	3	3
Duration (min)	4.10	3.25	4.51	5.28	5.28
FOV[c] (mm)	170	180	130	180	180
Matrix size	304 x 228	256 x 205	288 x 202	304 x 212	256 x 205
Reconstruction matrix	512	512	512	512	512

[a]: *the non-optional abduction external rotation view improves visualization of rotator cuff partial-thickness tears and the anterior labroligamentous complex*

[b]: *number of excitations*

[c]: *field of view*

shoulder disease. During surgery, the shoulder was structurally inspected according to a seven-item scoring list rating presence and severity of any lesion. The assessed disease entities were (Figures 1-5): fracture of the greater humeral tuberosity (not present / present), cuff lesion (not present / partial tear / full-thickness tear), Hill-Sachs lesion (not present / degeneration / Hill-Sachs), joint capsule lesion (not present / redundancy / anterior stripping or tear), superior labral tear from anterior to posterior (SLAP) lesion (not present / present), bony Bankart lesion (not present / degeneration / bony Bankart), and classic Bankart lesion (not present / degeneration / Bankart).

Subsequently, all acquired MRA's were prospectively evaluated by six radiologists according to the same seven-item scoring list to evaluate the influence of the radiologist's experience level, assessment in consensus, and ABER on diagnostic performance in three different sessions. Two sessions were individually (independent assessment of 58 MRA's without and independent assessment of 51 MRA's with ABER) and one was per team (consensus assessment of 58 MRA's). For this purpose three teams of two radiologists were assembled, each paired according to their estimated level of experience at the beginning of the study. The first team (T1) comprised the two most experienced dedicated musculoskeletal radiologists R1 and R2 (13 and 1 years of musculoskeletal experience,

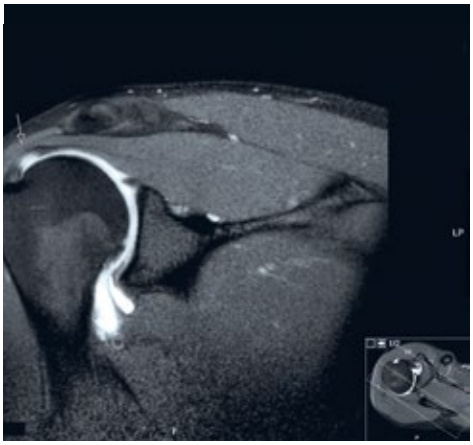

Figure 1: Example of a partial cuff tear (arrow) (oblique coronal orientation, T1-weighted): defect of the m. supraspinatus on the articular side

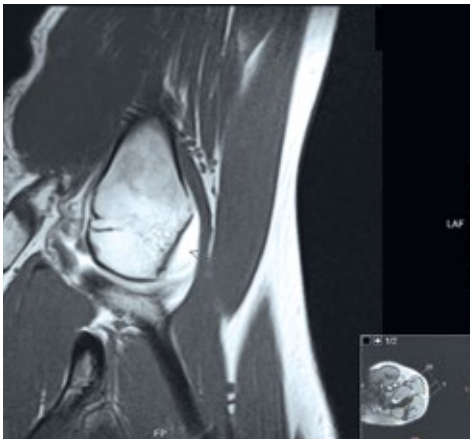

Figure 2: Example of a Hill-Sachs lesion (arrow) (ABER position orientation, T1-weighted): posterior-lateral bone defect of the humeral head

respectively). The second team (T2) consisted of two very experienced general radiologists R3 and R4 (17 and 15 years, respectively). R3 was, however, more musculoskeletal interested than R4. The third team (T3) was the least experienced of all [a first year general radiologist (R5) and a radiology resident (R6)].

The three sessions were separated by a 2-month interval (to avoid MRA recognition) and each session was divided in four meetings (to avoid fatigue) in which MRA's were randomly presented. In 45 MRA's the surgical report served as the reference standard. The spectrum of assessed lesions ranged from absence of abnormalities (Fig 6) to the presence of several lesions. All radiologists were blinded to patient data, clinical history, surgical report, and the MRA findings of other radiologists. However, both radiologists and orthopaedic surgeons were provided with a list of definitions of all lesion types to create uniformity.

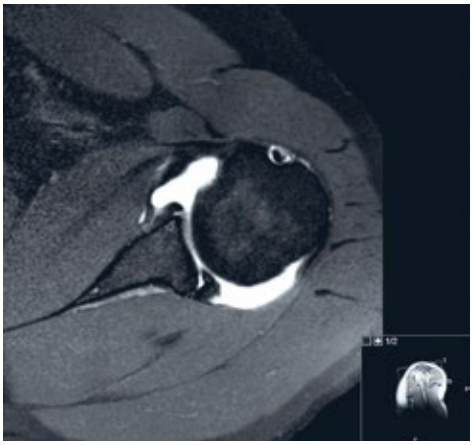

Figure 3: Example of an extensive labrum lesion (axial orientation, T1-weighted): avulsion of the anterior-inferior labrum from the anterior-inferior rim of the glenoid (long arrow) extended to posterior-inferior (short arrow)

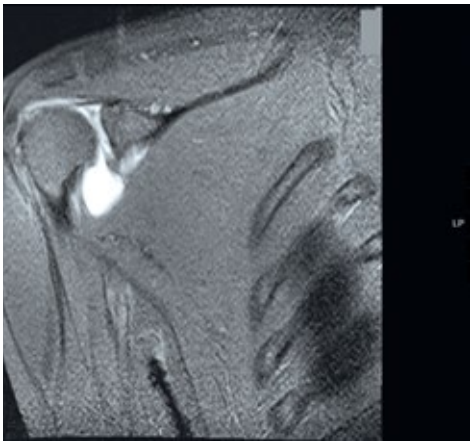

Figure 4: Example of a SLAP lesion type 5 (arrow) (oblique coronal orientation, T1-weighted): labrum lesion continues from anterior-inferior to anterior-superior and superior-posterior

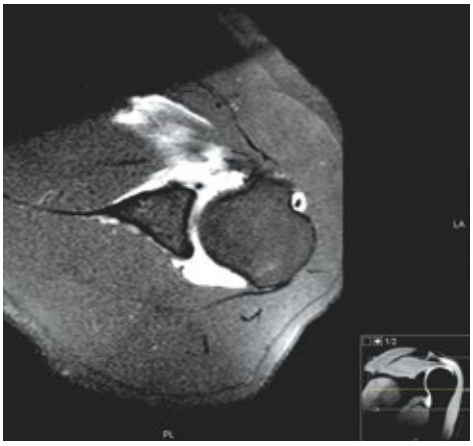

Figure 5: Example of a joint capsule lesion (arrow) (axial orientation, T1-weighted): stripping of the anterior capsulo-ligamentous complex from the anterior rim of the glenoid

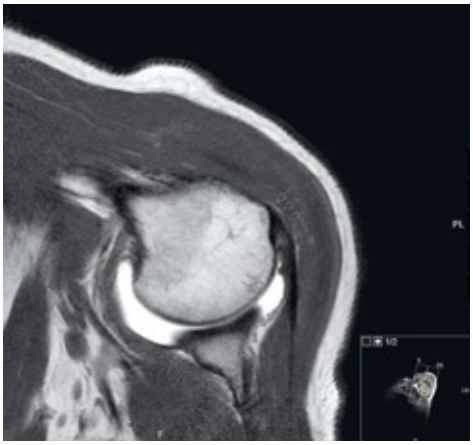

Figure 6: Example of three normal structures (ABER position orientation, T1-weighted): normal labral attachment on the anterior-inferior glenoid (single arrow), a normal humeral head contour (double arrow) and a part of the normal footprint attachment of the supraspinatus tendon (triple arrow)

Statistical analysis

MRA reproducibility between the individual radiologists and between teams was expressed as kappa coefficients and absolute percentage of agreement per different lesion type and overall per seven lesion types. Kappa values of 0.00-0.20 indicate poor agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.00 almost perfect agreement. [42]

MRA accuracy of the individual radiologists and of the teams was expressed in sensitivity, specificity and percentage of correct diagnosis values, as confirmed by arthroscopy, per different lesion type and overall per seven lesion types. Influence of experience, assessment in consensus and ABER on diagnostic MRA reproducibility and accuracy were evaluated by comparing the percentage of agreement or percentage of correct diagnosis of the radiologist and teams per different lesion type and overall per seven lesion types. Significance of differences was tested using the McNemar tests for paired proportions. [43] No adjustment for multiple testing was made. A p-value of ≤ 0.05 was considered statistically significant.

The seven assessed lesion types were dichotomized to construct 2 x 2 tables to facilitate calculations. All statistical analyses were conducted with SPSS 15.0 software. Post hoc sample size calculations [3] indicated that 40 MRA's were required to statistically detect with a power of 80% and a type 1 error of 5%, a 15% difference in correctly diagnosed lesions between the radiologists when the proportion of discordant pairs is conservatively assumed 0.65 and the method of analysis is McNemar test of equality of paired proportions.

Results

The final study population consisted of 58 patients (39 men, 19 women), mean age at MRA was 29.6 years (range 17-57 years). Fifty patients were referred for surgical stabilization after MRA, five patients refused surgery resulting in 45 surgery-confirmed MRA diagnoses (31 men and 14 women). The mean time from MRA to surgery was 4.9 months (range 1-24 months). To the authors' knowledge, none of the patients had a recurrent shoulder dislocation in this period. There were no adverse events reported after MRA and arthroscopy.

In 48 patients all MRA sequences were performed, in 10 cases one MRA sequence was missing due to poor execution of the study protocol (seven ABER, one sagittal, and two T1-weighted coronal oblique sequences). Two MRA examinations were affected by suboptimal administration of contrast agent and seven ABER sequences were affected by motion artefacts or malposition of the affected arm. Radiological assessments were performed over a period of 3.5 years, between 2008 and 2011. All data of the individual radiologist, teams and arthroscopic findings of the orthopaedic surgeons are complete.

MRA reproducibility between the individual radiologists and teams (Table 2a and 2b)

The overall kappa between the individual radiologists was poor (k = 0.17), to moderate (k = 0.48) with ABER, and fair (k = 0.30) to moderate (k = 0.44) without ABER. The overall absolute percentage of agreement between the most experienced dedicated musculoskeletal radiologists (R1-R2) was higher than those between the experienced general (R3-R4) and

Table 2a: Reproducibility of MRA diagnosis per lesion type and overall per 7 lesion types (N=58, 406 pair-wise ratings) between individual radiologists without ABER

	R1/R2	***R3/R4***	***R5/R6***
Cuff lesion‡	0.63/86.2	0.41/77.6	0.40/72.4
Humeral head lesion‡	0.62/81.0	0.38/69.0	0.48/79.3
Anterior inferior glenoid lesion‡	NA/89.7	0.13/84.5	-0.08/81.0
Greater humeral tuberosity fractures‡	NA/98.3	1.00/100.0	1.00/100.0
Anterior inferior labrum lesion‡	0.23/58.6	0.08/44.8	0.32/62.1
SLAP lesion‡	0.28/65.5	0.37/74.1	-0.05/50.0
Joint capsule lesion‡	0.09/62.1	0.10/62.1	0.09/50.0
Overall absolute percentage of agreement	**77.3**	**73.2**	**70.7**
Overall kappa (95% CI)	**0.44(0.35-0.54)**	**0.30(0.20-0.40)**	**0.31(0.22-0.40)**

NA: not applicable (calculation of kappa value not possible due to empty cells in 2x2 table), MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, R1/2/3/4/5/6: radiologist 1/2/3/4/5/6, ‡: κ-coefficient/absolute percentage of agreement (%)

Table 2b: Reproducibility of MRA diagnosis between individual radiologists with ABER (N=51) and between teams (N=58, 7 without ABER), per lesion type and overall per 7 lesion types (N=51, 357 pair-wise ratings or in case of N=58, 406 pair-wise ratings)

	R1/R2	***R3/R4***	***R5/R6***	***T1/T2***	***T2/T3***
Cuff lesion‡	0.70/88.2	0.19/72.5	0.31/70.6	0.22/75.9	0.53/82.8
Humeral head lesion‡	0.60/80.4	0.27/62.7	0.42/74.5	0.82/91.4	0.54/75.9
Anterior inferior glenoid lesion‡	0.37/94.1	NA/78.4	0.26/84.3	0.40/91.4	0.21/89.7
Greater humeral tuberosity fractures‡	1.00/100.0	1.00/100.0	1.00/100.0	1.00/100.0	1.00/100.0
Anterior inferior labrum lesion‡	0.23/52.9	0.03/39.2	0.20/58.8	0.40/72.4	0.08/50.0
SLAP lesion‡	0.05/58.8	0.29/74.5	-0.05/56.9	0.05/70.7	-0.06/72.4
Joint capsule lesion‡	0.34/72.5	-0.04/66.7	0.01/51.0	-0.07/70.7	0.08/55.2
Overall absolute percentage of agreement	**78.2**	**70.6**	**70.9**	**81.8**	**75.1**
Overall kappa (95% CI)	**0.48 (0.38-0.57)**	**0.17 (0.06-0.28)**	**0.25 (0.14-0.36)**	**0.53 (0.44-0.63)**	**0.33 (0.23-0.44)**

NA: not applicable (calculation of kappa value not possible due to empty cells in 2x2 table, MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, R1/2/3/4/5/6: radiologist 1/2/3/4/5/6, T1/2/3: team 1/2/3, ‡: κ-coefficients/ absolute percentage of agreement (%)

least experienced (R5-R6) radiologists with ABER (78.2% versus 70.6% and 70.9%) and without ABER (77.3% versus 73.2% and 70.7%). The same pattern of overall diagnostic reproducibility was established between the consensus assessments of the teams. The overall kappa was fair (k = 0.33) to moderate (k = 0.53), whereas the overall absolute percentage of agreement of the most experienced teams (T1-T2) was exceedingly higher than that of the lesser experienced teams (T2-T3), 81.8% versus 75.1%, respectively.

MRA accuracy of the individual radiologists and teams (Table 3)

The overall sensitivity of the individual radiologists ranged from 33.6-58.9% without ABER and from 30.6-63.5% with ABER. Overall specificity was higher and ranged from 73.6-89.9% without ABER and from 76.8-89.0% with ABER. The overall percentage of correct diagnosis of the three most experienced (musculoskeletal) radiologists (R1, R2, and R3) was higher than those of the lesser experienced radiologists (R4, R5, and R6) with ABER (74.4%, 74.8%, 75.2% versus 68.8%, 71.8%, 68.4) and without ABER (73.3%, 71.4%, 74.9% versus 66%, 70.2%, 68.6%). The same pattern of overall diagnostic accuracy was established by the teams. Overall sensitivity was low and ranged from 43.0-57.9%. Overall specificity was higher and ranged from 85.1-88.5% and the overall percentage of

Table 3: Accuracy (% Se, SP or Acc) of MRA diagnoses of the individual radiologists (first % in each cell is without ABER (N=45), second % is with ABER (N=38)) and of the teams (N=45, 7 without ABER), per lesion type and overall per 7 lesion types (N=45, 315 pair-wise ratings or in case of N=38, 266 pair-wise ratings)

		Radiologist 1			***Radiologist 2***	
	Sens%	Spec%	Acc%	Sens%	Spec%	Acc%
Cuff lesion	80.0/80.0	90.0/90.9	88.9/89.5	80.0/80.0	85.0/78.8	84.4/78.9
Humeral head lesion	64.3/68.2	58.8/56.3	62.2/63.2	75.0/77.7	52.9/43.8	66.7/63.2
Anterior inferior glenoid lesion	0.0/0.0	100.0/94.6	93.3/92.1	33.3/100.0	95.2/97.3	91.1/97.4
Greater humeral tuberosity fractures	NA/NA	97.8/100.0	97.8/100.0	NA/NA	100.0/100.0	100.0/100.0
Anterior inferior labrum lesion	34.2/41.9	100.0/85.7	44.4/50.0	60.5/80.6	60.5/42.9	60.0/73.7
SLAP lesion	0.0/50.0	83.7/86.1	80.0/84.2	100.0/50.0	58.1/63.9	60.0/63.2
Joint capsule lesion	29.0/25.0	85.7/71.4	46.7/42.1	22.6/25.0	71.4/85.7	37.8/47.4
Overall Se, Sp and Acc	**41.1/45.9**	**89.9/87.8**	**73.3/74.4**	**54.2/63.5**	**80.3/80.1**	**71.4/74.8**
		Radiologist 3			***Radiologist 4***	
	Sens%	Spec%	Acc%	Sens%	Spec%	Acc%
Cuff lesion	80.0/80.0	90.0/90.9	88.9/89.5	40.0/40.0	77.5/81.8	73.3/76.3
Humeral head lesion	64.3/68.2	52.9/43.8	60.0/57.9	35.7/40.9	70.6/62.5	48.9/50.0
Anterior inferior glenoid lesion	66.7/0.0	85.7/78.4	84.4/76.3	33.3/0.0	97.6/100.0	93.3/97.4
Greater humeral tuberosity fractures	NA/NA	100.0/100.0	100.0/100.0	NA/NA	100.0/100.0	100.0/100.0
Anterior inferior labrum lesion	84.2/90.3	57.1/57.1	80.0/84.2	28.9/19.4	85.7/100.0	37.8/34.2
SLAP lesion	0.0/0.0	76.7/88.9	73.3/84.2	0.0/50.0	67.4/72.2	64.4/71.1
Joint capsule lesion	16.1/0.0	85.7/92.9	37.8/34.2	38.7/33.3	57.1/85.7	44.4/52.6
Overall Se, Sp and Acc	**57.0/55.3**	**84.4/84.5**	**74.9/75.2**	**33.6/30.6**	**82.7/86.7**	**66.0/68.8**
		Radiologist 5			***Radiologist 6***	
	Sens%	Spec%	Acc%	Sens%	Spec%	Acc%
Cuff lesion	80.0/80.0	87.5/81.8	86.7/81.6	80.0/20.0	67.5/66.7	68.9/60.5
Humeral head lesion	39.3/40.9	76.5/75.0	53.3/55.3	32.1/36.4	76.5/68.8	48.9/50.0
Anterior inferior glenoid lesion	0.0/0.0	95.2/94.6	88.9/92.1	0.0/100.0	88.1/89.2	82.2/89.5
Greater humeral tuberosity fractures	NA/NA	100.0/100.0	100.0/100.0	NA/NA	100.0/100.0	100.0/100.0
Anterior inferior labrum lesion	35.7/29.0	71.4/71.4	33.3/36.8	65.8/64.5	42.9/85.7	62.2/68.4
SLAP lesion	0.0/50.0	86.0/86.1	82.2/84.2	0.0/100.0	55.8/72.2	53.3/73.7
Joint capsule lesion	35.5/29.2	71.4/92.9	46.7/52.6	80.6/45.8	28.6/21.4	64.4/36.8
Overall Se, Sp and Acc	**33.6/35.3**	**88.9/89.0**	**70.2/71.8**	**58.9/50.6**	**73.6/76.8**	**68.6/68.4**

Table 3: Continued

	Team 1			*Team 2*		
	Sens%	Spec%	Acc%	Sens%	Spec%	Acc%
Cuff lesion	80.0	92.5	91.1	60.0	85.0	82.2
Humeral head lesion	78.6	41.2	64.4	71.4	41.2	60.0
Anterior inferior glenoid lesion	0.0	95.2	88.9	0.0	90.5	84.4
Greater humeral tuberosity fractures	NA	100.0	100.0	NA	100.0	100.0
Anterior inferior labrum lesion	73.7	57.1	71.1	86.8	85.7	86.7
SLAP lesion	100.0	79.1	80.0	0.0	95.3	91.1
Joint capsule lesion	16.1	78.6	35.6	19.4	92.9	42.2
Overall Se, Sp and Acc	**57.0**	**85.6**	**75.9**	**57.9**	**88.5**	**78.1**

	Team 3		
	Sens%	Spec%	Acc%
Cuff lesion	80.0	82.5	82.2
Humeral head lesion	39.3	64.7	48.9
Anterior inferior glenoid lesion	0.0	97.6	91.1
Greater humeral tuberosity fractures	NA	100.0	100.0
Anterior inferior labrum lesion	36.8	71.4	42.2
SLAP lesion	50.0	81.4	80.0
Joint capsule lesion	51.6	50.0	51.1
Overall Se, Sp and Acc	**43.0**	**85.1**	**70.8**

NA: not applicable, calculation not possible due to empty cells in 2 x 2 table, MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, Sens: sensitivity percentage, Spec: specificity percentage, Acc: percentage of correctly diagnosed lesions calculated as (true-positive + true-negative) / total patients in 2x2 table

correct diagnosis of the two most experienced teams (T1,T2) were much better than that of the least experienced team (T3), 75.9% and 78.1% versus 70.8%, respectively.

The influence of experience level on MRA reproducibility (Table 4a and 4b)

The most experienced dedicated musculoskeletal radiologists (R1-R2) agreed significantly more, per seven lesion types, than the experienced general radiologists R3-R4 (with ABER; p = 0.014) and the least experienced radiologists R5-R6 (with ABER; p = 0.018 and without ABER; p = 0.030). Per lesion type, a significant difference in percentage agreement was found, when comparing R1-R2 versus R3-R4 or R5-R6, in case of cuff (p = 0.039, p = 0.035), anterior inferior glenoid (p = 0.039), and joint capsule lesions (p = 0.043). The same pattern was established between the paired teams. The more experienced teams (T1-T2) agreed significantly more (p = 0.007) than the lesser experienced teams (T2-T3). Per lesion type, superior reproducibility between T1-T2 and T2-T3 was found in case of humeral head (p = 0.049) and anterior inferior labrum lesions (p = 0.004).
The influence of experience level on MRA accuracy (Table 5)

Table 4a: Influence of experience level on reproducibility of MRA diagnosis. Differences in percentages of agreement between the paired individual radiologists, without ABER, are tested using the McNemar test for paired proportions. P-values per lesion type and overall per 7 lesion types (N=58, 406 pair-wise ratings)

	R1R2/R3R4	***R1R2/R5R6***	***R3R4/R5R6***
Cuff lesion	0.267	0.115	0.690
Humeral head lesion	0.167	1.000	0.307
Anterior inferior glenoid lesion	0.508	0.302	0.804
Greater humeral tuberosity fractures	NA	NA	NA
Anterior inferior labrum lesion	0.169	0.851	0.110
SLAP lesion	0.405	0.136	0.016*
Joint capsule lesion	1.000	0.281	0.281
Overall absolute percentage of agreement	**0.135**	**0.030***	**0.474**
Superior overall agreement	**R1R2**	**R1R2‡**	**R3R4**

*NA: not applicable (calculation not possible due to empty cells in 2 x 2 table), ‡: significant superior overall agreement, MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, R1/2/3/4/5/6: radiologist 1/2/3/4/5/6, *: significant difference at p-level <0.05*

Table 4b: Influence of experience level on reproducibility of MRA diagnosis. Differences in percentages of agreement between the paired individual radiologists with ABER (N=51) and between the paired teams (N=58, 7 without ABER) are tested using the McNemar test for paired proportions. P-values per lesion type and overall per 7 lesion types (N=51, 357 pair-wise ratings or in case of N=58, 406 pair-wise ratings)

	R1R2/R3R4	***R1R2/R5R6***	***R3R4/R5R6***	***T1T2/T2T3***
Cuff lesion	0.039*	0.035*	1.000	0.344
Humeral head lesion	0.064	0.629	0.307	0.049*
Anterior inferior glenoid lesion	0.039*	0.063	0.607	1.000
Greater humeral tuberosity fractures	NA	NA	NA	NA
Anterior inferior labrum lesion	0.248	0.701	0.052	0.052
SLAP lesion	0.115	1.000	0.093	1.000
Joint capsule lesion	0.678	0.043*	0.152	0.122
Overall absolute percentage of agreement	**0.014***	**0.018***	**1.000**	**0.007***
Superior overall agreement	**R1R2‡**	**R1R2‡**	**Equal**	**T1T2‡**

*NA: not applicable (calculation not possible due to empty cells in 2x2 table, MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, R1/2/3/4/5/6: radiologist 1/2/3/4/5/6, T1/2/3: team 1/2/3, *: significant difference at p-level <0.05), ‡: significant superior overall agreement*

Although the three most experienced (musculoskeletal) radiologists (R1, R2, and R3) were consistently more accurate than the lesser experienced radiologists (R4, R5, and R6) with and without ABER, the overall percentage of correct diagnosis was only significant different between R1-R4 (p = 0.012) and R3-R4 (p = 0.003), without ABER. Per lesion type, superior significant accuracy by means of experience level was especially found in case of cuff and anterior inferior labrum lesions. The same pattern was encountered by the teams. Although the two most experienced teams (T1, T2) were much more accurate than the least experienced team (T3), the overall percentage of correct diagnosis was only significant different between T2 and T3 (p = 0.014). Per lesion type, superior significant accuracy by means of experience level was found in case of anterior inferior labrum lesions.

Table 5: Influence of experience level on accuracy of MRA diagnosis. Differences in percentages correctly diagnosed, between the individual radiologists (without ABER N=45, and with ABER N=38) and between the teams (N=45, 7 without ABER), are tested using the McNemar test for paired proportions. P-values per lesion type and overall per 7 lesion types (N=45, 315 pair-wise ratings or in case of N=38, 266 pair-wise ratings)

	R1/R4	***R1/R5***	***R1/R6***	***R2/R4***
Cuff lesion	0.039*/0.125	1.000/0.453	0.012*/0.003*	0.227/1.000
Humeral head lesion	0.210/0.267	0.424/0.508	0.210/0.302	0.077/0.302
Anterior inferior glenoid lesion	1.000/0.500	0.344/1.000	0.031*/1.000	1.000/1.000
Greater humeral tuberosity fractures	NA/NA	NA/NA	NA/NA	NA/NA
Anterior inferior labrum lesion	0.629/0.146	0.227/0.227	0.134/0.092	0.041*/0.003*
SLAP lesion	0.092/0.227	1.000/1.000	0.012*/0.424	0.815/0.508
Joint capsule lesion	1.000/0.454	1.000/0.424	0.169/0.824	0.629/0.791
Overall, per 7 lesion types	**0.012*/0.072**	**0.193/0.371**	**0.163/0.093**	**0.086/0.072**
Which radiologist / team scores best?	**R1‡/R1**	**R1/R1**	**R1/R1**	**R2/R2**
	R2/R5	***R2/R6***	***R3/R4***	***R3/R5***
Cuff lesion	1.000/1.000	0.065/0.065	0.039*/0.180	1.000/0.453
Humeral head lesion	0.210/0.581	0.077/0.332	0.302/0.581	0.607/1.000
Anterior inferior glenoid lesion	1.000/0.500	0.289/0.250	0.289/0.008*	0.727/0.031*
Greater humeral tuberosity fractures	NA/NA	NA/NA	NA/NA	NA/NA
Anterior inferior labrum lesion	0.004*/0.001*	1.000/0.804	<0.001*/<0.001*	<0.001*/<0.001*
SLAP lesion	0.021*/0.021*	0.664/0.481	0.344/0.125	0.424/1.000
Joint capsule lesion	0.388/0.791	0.023*/0.523	0.629/0.065	0.454/0.039*
Overall, per 7 lesion types	**0.720/0.389**	**0.412/0.086**	**0.003*/0.060**	**0.124/0.321**
Which radiologist / team scores best?	**R2/R2**	**R2/R2**	**R3‡/R3**	**R3/R3**
	R3/R6	***T1/T2***	***T1/T3***	***T2/T3***
Cuff lesion	0.012*/0.003*	0.344	0.289	1.000
Humeral head lesion	0.359/0.607	0.625	0.118	0.267
Anterior inferior glenoid lesion	1.000/0.180	0.625	1.000	0.250
Greater humeral tuberosity fractures	NA/NA	NA	NA	NA
Anterior inferior labrum lesion	0.096/0.146	0.092	0.001*	<0.001*
SLAP lesion	0.035*/0.344	0.227	1.000	0.227
Joint capsule lesion	0.043*/1.000	0.581	0.189	0.523
Overall, per 7 lesion types	**0.064/0.057**	**0.419**	**0.072**	**0.014***
Which radiologist / team scores best?	**R3/R3**	**T2**	**T1**	**T2‡**

*NA: not applicable, calculation not possible due to empty cells in 2 x 2 table, MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, R1/2/3/4/5/6: radiologist 1/2/3/4/5/6, T1/2/3: team 1/2/3, *: significant difference at p-level <0.05, ‡: significant superior overall agreement*

The influence of consensus assessment on MRA reproducibility (Table 6)

All the paired teams (agreement between consensus assessments) agreed more per seven lesion types than the individual radiologists of each team alone (agreement between individual assessments) with or without ABER. The difference in percentage of agreement was, however, only significant between T1-T2 and their lesser experienced radiologists pair R3-R4 ($p < 0.001$ and $p = 0.001$, respectively). Per lesion type, a significant superior reproducibility through consensus assessment was especially found in case of humeral head and anterior inferior labrum lesions.

Table 6: Influence of consensus assessment on reproducibility of MRA diagnosis. Differences in percentages of agreement between the paired teams and their consisting paired individual radiologists, without (N=58) and with ABER (N=51), are tested using the McNemar test for paired proportions. P-values per lesion type and overall per 7 lesion types (N=51, 357 pair-wise ratings or in case of N=58, 406 pair-wise ratings)

	T1T2/R1R2	***T1T2/R3R4***	***T2T3/R3R4***	***T2T3/R5R6***
Cuff lesion	0.210/0.238	1.000/0.815	0.607/0.424	0.210/0.302
Humeral head lesion	0.180/0.146	0.004*/0.001*	0.481/0.263	0.815/1.000
Anterior inferior glenoid lesion	1.000/1.000	0.388/0.092	0.549/0.267	0.267/0.754
Greater humeral tuberosity fractures	NA/NA	NA/NA	NA/NA	NA/NA
Anterior inferior labrum lesion	0.169/0.013*	0.004*/0.001*	0.711/0.096	0.230/0.824
SLAP lesion	0.664/0.180	0.791/0.804	1.000/1.000	0.019*/0.115
Joint capsule lesion	0.359/1.000	0.383/0.824	0.585/0.405	0.711/0.664
Overall absolute percentage of agreement (N=406)/(N=357)	**0.101/0.128**	**0.001*/<0.001***	**0.536/0.122**	**0.133/0.137**
Superior overall agreement	**T1T2**	**T1T2‡**	**T2T3**	**T2T3**

*NA: not applicable (calculation not possible due to empty cells in 2 x 2 table), MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, R1/2/3/4/5/6: radiologist 1/2/3/4/5/6, T1/2/3: team 1/2/3, *: significant difference at p-level <0.05, ‡: significant superior overall agreement*

The influence of consensus assessment on MRA accuracy (Table 7)

All the teams (consensus assessment) were more accurate per seven lesion types than their consisting individual radiologists (individual assessment) with or without ABER. The difference in percentage of correct diagnosis was, however, only significant between the T2 and R3 ($p = 0.050$ with ABER) or R4 ($p = 0.001$ with and $p \leq 0.001$ without ABER). Per lesion type, a significant superior accuracy through consensus assessment was especially found in case of SLAP and anterior-inferior labrum lesions.

Table 7: Influence of consensus assessment on accuracy on MRA diagnosis. Differences in percentages correctly diagnosed, between the teams and their consisting individual radiologists, without (N=45) and with ABER (N=38) are tested using the McNemar test for paired proportions. P-values per lesion type and overall per 7 lesion types (N=45, 315 pair-wise ratings or in case of N=38, 266 pair-wise ratings)

	T1/R1	***T1/R2***	***T2/R3***
Cuff lesion	1.000/1.000	0.508/0.125	0.453/0.250
Humeral head lesion	1.000/1.000	1.000/1.000	1.000/1.000
Anterior inferior glenoid lesion	0.500/1.000	1.000/1.000	1.000/0.289
Greater humeral tuberosity fractures	NA/NA	NA/NA	NA/NA
Anterior inferior labrum lesion	0.017[a]/0.022[a]	0.227/1.000	0.508/0.625
SLAP lesion	1.000/0.727	0.012[a]/0.070	0.021[a]/0.375
Joint capsule lesion	0.267/1.000	1.000/0.581	0.754/0.125
Overall, per 7 lesion types	**0.358/0.291**	**0.081/0.349**	**0.220/0.050[a]**
Which radiologist / team scores best?	**T1**	**T1**	**T2∞**
	T2/R4	***T3/R5***	***T3/R6***
Cuff lesion	0.388/0.754	0.625/1.000	0.146/0.008[a]
Humeral head lesion	0.332/0.424	0.754/0.727	1.000/1.000
Anterior inferior glenoid lesion	0.125/0.125	1.000/1.000	0.219/0.625
Greater humeral tuberosity fractures	NA/NA	NA/NA	NA/NA
Anterior inferior labrum lesion	<0.001[a]/<0.001[a]	0.219/0.289	0.049[b]/0.039[b]
SLAP lesion	0.002[a]/0.039[a]	1.000/0.754	0.008[a]/0.754
Joint capsule lesion	1.000/0.774	0.815/1.000	0.210/0.180
Overall, per 7 lesion types	**<0.001[a]/0.001[a]**	**0.888/1.000**	**0.888/1.000**
Which radiologist / team scores best?	**T2‡**	**T3**	**T3**

NA: not applicable, calculation not possible due to empty cells in 2 x 2 table, MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, R1/2/3/4/5/6: radiologist 1/2/3/4/5/6, T1/2/3: team 1/2/3, [a]: significant difference at p-level <0.05 in favor of consensus assessment, [b]: significant difference at p-level <0.05 in favor of the individual radiologist, ∞: significant superior overall agreement with ABER, ‡: significant superior overall agreement

The influence of ABER on MRA reproducibility (Table 8) and accuracy (Table 9)

There were no significant differences in percentage of agreement per lesion or overall per seven lesions between the paired individual radiologists with or without ABER. No overall significant differences in percentage of correct diagnoses of the individual radiologists were found with or without ABER. The utilization of ABER did not increase diagnostic MRA reproducibility or accuracy.

Table 8: Influence of ABER on reproducibility of MRA diagnosis. Differences in percentages of agreement between the paired individual radiologists, with and without ABER, are tested using the McNemar test for paired proportions. P-values per lesion type and overall per 7 lesion types (N=51, 357 pair-wise ratings)

	R1R2/R1R2	***R3R4/R3R4***	***R5R6/R5R6***
Cuff lesion	0.727	0.322	1.000
Humeral head lesion	0.508	0.424	0.791
Anterior inferior glenoid lesion	0.687	0.607	0.508
Greater humeral tuberosity fractures	NA	NA	NA
Anterior inferior labrum lesion	0.405	0.541	1.000
SLAP lesion	0.815	0.791	0.442
Joint capsule lesion	0.267	0.541	1.000
Overall absolute percentage of agreement (N=406)/(N=357)	**1.000**	**0.387**	**0.769**
Superior overall agreement	**Equal**	**R3R4, without ABER**	**R5R6, with ABER**

NA: not applicable (calculation not possible due to empty cells in 2 x 2 table), MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, R1/2/3/4/5/6: radiologist 1/2/3/4/5/6

Table 9: Influence of ABER on accuracy of MRA diagnosis. Differences in percentages correctly diagnosed, between the individual radiologist's assessment without and with ABER, are tested using the McNemar test for paired proportions. P-values per lesion type and overall per 7 lesion types (N=38, 266 pair-wise ratings)

	R1	***R2***	***R3***	***R4***	***R5***	**R6**
Cuff lesion	1.000	0.500	1.000	1.000	1.000	0.304
Humeral head lesion	0.688	1.000	1.000	1.000	1.000	1.000
Anterior inferior glenoid lesion	0.500	0.625	0.581	1.000	1.000	0.688
Greater humeral tuberosity fractures	NA	NA	NA	NA	NA	NA
Anterior inferior labrum lesion	1.000	0.180	0.688	0.549	1.000	1.000
SLAP lesion	0.688	0.754	0.125	0.344	1.000	0.134
Joint capsule lesion	0.453	0.754	0.375	0.791	1.000	0.031*
Overall, per 7 lesion types	**0.736**	**0.174**	**1.000**	**0.890**	**1.000**	**0.911**
What scores best? With or without ABER	**Without**	**With**	**With**	**With**	**Without**	**Without**

*NA: not applicable, calculation not possible due to empty cells in 2 x 2 table, MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, R1/2/3/4/5/6: radiologist 1/2/3/4/5/6, *: significant difference at p-level <0.05*

Discussion

The most important finding of the present study is that the addition of ABER does not significantly improve the diagnostic performance of high-field MRA in patients with TASI. However, the radiologist's experience level and assessment in consensus do contribute to diagnostic reproducibility and accuracy improvement.

For the radiologists and teams combined, the overall kappa ranged from poor (k = 0.17) to moderate (k = 0.53), sensitivity from 30.6-63.5%, specificity from 73.6-89.9%, and accuracy from 66.0-78.1%. Although the overall kappa's and specificities were roughly in line with earlier literature (k = -0.03-0.84, 85-100%), the overall sensitivities were much lower than previously reported by other authors (79-100%). [6,12,19,20,25,26,28-40] The overall accuracy rates were, however, in perfect concurrence with a former retrospective study [3], where 61 MRA's were assessed for the same seven lesion types resulting in a sensitivity of 50.0-60.6%, a specificity of 66.8-87.9%, and an accuracy of 65.4-78.9%.

The level of experience of the participating radiologists was expected to positively influence the diagnostic MRA reproducibility and accuracy achieved: results of the present study confirm this hypothesis. Overall, the most experienced radiologists (R1-R2) and teams (T1-T2) agreed significantly more about the presence of seven lesion types while assessing TASI-related MRA's than the lesser experienced radiologists R3-R4 ($p = 0.014$), R5-R6 ($p = 0.018$ / $p = 0.030$), and teams T2-T3 ($p = 0.007$). Although a significant overall higher accuracy could only be established between R1-R4 ($p = 0.012$), R3-R4 ($p = 0.03$), and T2-T3 ($p = 0.014$), a firm trend of higher accuracy by means of experience level was detected. The more experienced radiologists (R1, R2, R3) and teams (T1, T2) were consistently more accurate than the lesser experienced radiologist (R4, R5, R6) and team (T3). This positive influence of experience level on diagnostic accuracy is supported by former literature. A retrospective study by Theodoropoulos et al. [36] compared the diagnostic MRA performance of general radiologists with musculoskeletal radiologists in 250 labral, rotator cuff, biceps, and Hill-Sachs lesions. The musculoskeletal radiologists were significant more sensitive (75-87%) than the general radiologists (20-60%). A second, one-lesion, prospective study by Reuss et al. [41] compared the pooled sensitivity of 28 community radiologists (51%) with two academic musculoskeletal radiologists (77%) diagnosing SLAP lesions on 83 MRA's and concluded that one of the two academic radiologists was significantly more sensitive ($p = 0.0063$).

A reason that the present accuracy results (trend) were not totally in line with the significant reproducibility results could be that there is no standard personal feedback in the Rijnstate Hospital, considering the MRA findings of the radiologists after surgical stabilization. This might limit their professional development and consequently a higher degree of specialized experience might not lead to a significant higher diagnostic accuracy. Furthermore, the total MRA assessment period extended over 3.5 years. Considering the lack of feedback, this might have caused a steeper learning curve in the lesser experienced

radiologists, reducing the total influence of experience level on diagnostic accuracy. Nevertheless, it is advisable to leave MRA assessments in experienced hands as systematic accurate MRA diagnosis of each separate radiologist is only achievable through high reproducibility.

The results of the present study also confirmed the hypothesis that assessment in consensus would positively influence the diagnostic reproducibility and accuracy of MRA. Although significant reproducibility or accuracy improvement through consensus assessment could only be demonstrated between T1-T2 and radiologists R3-R4 ($p < 0.001$ / $p = 0.001$) or between T2 and the individual assessments of R3 ($p < 0.001$ / $p = 0.001$) and R4 ($p = 0.050$), overall agreement and accuracy of the consensus assessments were systematically higher than the individual assessments, indicating a firm trend. Apparently, assessment in consensus complements the specific knowledge of the individual radiologists. Therefore, it is essential that even very experienced and specialized musculoskeletal radiologists should consult each other and discuss TASI-associated lesions that are difficult to diagnose at MRA in order to achieve a reproducible and accurate diagnosis.

To the authors' surprise, the addition of ABER did not significantly improve overall reproducibility and accuracy, as previous studies had described superior results assessing the rotator cuff and the anterior labroligamentous complex. [2,14,23,34,38] The present findings were, however, in accordance with two large studies conducted by Schreinemachers et al. [25,26] and one large study conducted by Tian et al. [34] Schreinemachers et al. assessed 250 MRA's for partial-thickness supraspinatus tears and anteroinferior labroligamentous lesions with and without ABER in a large general hospital. Their interobserver agreement ($k = 0.60$-0.67), sensitivity (52-70%), and specificity (91-95%) values for partial-thickness supraspinatus tears [25] did not significantly improve with the utilization of ABER ($k = 0.48$-0.65, 48-61%, and 80-94%, respectively). They found similar results in their second study, where the interobserver agreement ($k = 0.44$-0.62), sensitivity (89-96%), and specificity (84-89%) results for anteroinferior labroligamentous lesions [26] were not significantly improved either with the utilization of ABER ($k = 0.44$-0.56), 85-89%, and 82-91%, respectively). Tian et al. [34] assessed 229 MRA's for anteroinferior labroligamentous lesions in a large university hospital and could not found a significant difference in interobserver agreement between the neutral and ABER position either.

Although ABER is presumed to optimize visualization of TASI-associated lesions, it is also mentioned in former literature that approximately 8-20% of the patients are not able to tolerate the ABER position due to provocation of apprehension symptoms or pain, and that the arm positioned in ABER may be more susceptible to motion artefacts. [2,12,13,18,25,26,37] In fact, in the present study, 14% (7/51) of the MRA's were indeed affected by motion artefacts and in 12% of the cases (7/58) patients were not able to perform the ABER position. It is further mentioned that ABER is time consuming (adding at least 10 minutes to a routine MRA protocol) and there might be interpretation issues related to anatomical changes. [2,25,26]

Considering the lack of reproducibility and accuracy improvement, the results of the present study, the results of previous studies, the disadvantages of ABER utilization in general, and efficiency requirements in times of economic decline, the ABER sequence should not be incorporated in a MRA protocol.

A limitation of the present study is that the experience level of the six radiologists was ranked at the beginning of the study. These choices could have influenced the results, although the radiologists were classified to the best of the authors' abilities. Another limitation is the incomplete MRI sequence protocol in 10 out of 45 patients. A total of 10 sequences were missing (seven ABER, one sagittal, and two T1-weighted coronal oblique sequences). Although this could have negatively affected the results of the current study, the inability to assume the ABER position in seven patients is perfectly in line with the results of other authors. [2,12,13,18,25,26,37] Therefore, the three missing sequences out of 290 sequences (58 MRA's x 5 sequences) did not cause significant deviations of the present results.

The strengths of the present study are that it was the first prospective study evaluating the influence of experience level, consensus assessment, and ABER on diagnostic accuracy of MRA in a single institution using a consistent protocol with definitions of the assessed lesions to create uniformity among the reviewers.

Conclusion

In conclusion, the results of the present study show that the ABER sequence does not provide further additional diagnostic value. The experience level of the radiologists and assessment in consensus do, however, positively contribute to overall diagnostic performance of MRA in TASI-associated lesions. Standard personal MRA feedback after surgery and intensified intra-disciplinary consultation in cases that are difficult to diagnose at MRA should assure that radiologists attain an experienced-based steep learning curve in TASI lesion diagnosis. This will increase true professional development, increase diagnostic accuracy and will ultimately result in better healthcare.

Acknowledgements

The authors thank all participating radiologist of the Department of Radiology of the Rijnstate Hospital for the individual and consensus MRA assessments of TASI associated lesions.

References

1. Acid S, Le Corroller T, Aswad R, Pauly V, Champsaur P. Preoperative imaging of anterior shoulder instability: diagnostic effectiveness of MDCT arthrography and comparison with MR arthrography and arthroscopy. AJR Am J Roentgenol 2012;198:661-667.
2. Choi JA, Suh SI, Kim BH, Cha SH, Kim MG, Lee KY, et al. Comparison between conventional MR arthrography and abduction and external rotation MR arthrography in revealing tears of the antero-inferior glenoid labrum. Korean J Radiol 2001;2:216-221.
3. van Grinsven S, Kesselring FO, van Wassenaer-van Hall HN, Lindeboom R, Lucas C, van Loon CJ. MR arthrography of traumatic anterior shoulder lesions showed modest reproducibility and accuracy when evaluated under clinical circumstances. Arch Orthop Trauma Surg 2007;127:11-17.
4. van der Veen HC, Collins JP, Rijk PC. Value of magnetic resonance arthrography in post-traumatic anterior shoulder instability prior to arthroscopy: a prospective evaluation of MRA versus arthroscopy. Arch Orthop Trauma Surg 2012;132:371-375.
5. Waldt S, Burkart A, Imhoff AB, Bruegel M, Rummeny EJ, Woertler K. Anterior shoulder instability: accuracy of MR arthrography in the classification of anteroinferior labroligamentous injuries. Radiology 2005;237:578-583.
6. Halma JJ, Eshuis R, Krebbers YM, Weits T, de Gast A. Interdisciplinary inter-observer agreement and accuracy of MR imaging of the shoulder with arthroscopic correlation. Arch Orthop Trauma Surg 2012;132:311-320.
7. Jonas SC, Walton MJ, Sarangi PP. Is MRA an unnecessary expense in the management of a clinically unstable shoulder? Acta Orthop 2012;83:1-4.
8. Kirkley A, Litchfield R, Thain L, Spouge A. Agreement between magnetic resonance imaging and arthroscopic evaluation of the shoulder joint in primary anterior dislocation of the shoulder. Clin J Sport Med 2003;13:148-151.
9. Pandya NK, Colton A, Webner D, Sennett B, Huffman GR. Physical examination and magnetic resonance imaging in the diagnosis of superior labrum anteriorposterior lesions of the shoulder: a sensitivity analysis. Arthroscopy 2008;24:311-317.
10. Sanders TG, Morrison WB, Miller MD. Imaging techniques for the evaluation of glenohumeral instability. Am J Sports Med 2000;28:414-434.
11. Simao MN, Nogueira-Barbosa MH, Muglia VF, Barbieri CH. Anterior shoulder instability: correlation between magnetic resonance arthrography, ultrasound arthrography and intraoperative findings. Ultrasound Med Biol 2012;38:551-560.
12. Smith TO, Drew BT, Toms AP. A meta-analysis of the diagnostic test accuracy of MRA and MRI for the detection of glenoid labral injury. Arch Orthop Trauma Surg 2012;132:905-919.
13. Bencardino JT, Beltran J, Rosenberg ZS, Rokito A, Schmahmann S, Mota J, et al. Superior labrum anteriorposterior lesions: diagnosis with MR arthrography of the shoulder. Radiology 2000; 214:267-271.
14. Ito Y, Sakai T, Tomo H, Nakao Y, Inui K, Koike T, et al. Computerized assessment of Bankart lesions under tension with magnetic resonance arthrography. J Shoulder Elbow Surg 2005;14:247-251.
15. Amin MF, Youssef AO. The diagnostic value of magnetic resonance arthrography of the shoulder in detection and grading of SLAP lesions: comparison with arthroscopic findings. Eur J Radiol 2011;81:2343-2347.

16. Bak K, Wiesler ER, Poehling GG. Consensus statement on shoulder instability. Arthroscopy 2010;26:249-255.
17. Chandnani VP, Yeager TD, DeBerardino T, Christensen K, Gagliardi JA, Heitz DR, et al. Glenoid labral tears: prospective evaluation with MRI imaging, MR arthrography, and CT arthrography. AJR Am J Roentgenol 1993;161:1229-1235.
18. Cochet H, Couderc S, Pele E, Amoretti N, Moreau-Durieux MH, Hauger O, et al. Rotator cuff tears: should abduction and external rotation (ABER) positioning be performed before image acquisition? A CT arthrography study. Eur Radiol 2010;20:1234-1241.
19. Holzapfel K, Waldt S, Bruegel M, Paul J, Heinrich P, Imhoff AB, et al. Inter- and intraobserver variability of MR arthrography in the detection and classification of superior labral anterior posterior (SLAP) lesions: evaluation in 78 cases with arthroscopic correlation. Eur Radiol 2010;20:666-673.
20. Iqbal HJ, Rani S, Mahmood A, Brownson P, Aniq H. Diagnostic value of MR arthrogram in SLAP lesions of the shoulder. Surgeon 2010;8:303-309.
21. de Jesus JO, Parker L, Frangos AJ, Nazarian LN. Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. AJR Am J Roentgenol 2009;192:1701-1707.
22. Lee MJ, Motamedi K, Chow K, Seeger LL. Gradient-recalled echo sequences in direct shoulder MR arthrography for evaluating the labrum. Skeletal Radiol 2008;37:19-25.
23. Lee SY, Lee JK. Horizontal component of partial-thickness tears of rotator cuff: imaging characteristics and comparison of ABER view with oblique coronal view at MR arthrography initial results. Radiology 2002;224:470-476.
24. Matava MJ, Purcell DB, Rudzki JR. Partial-thickness rotator cuff tears. Am J Sports Med 2005;33:1405-1417.
25. Schreinemachers SA, van der Hulst VP, Willems WJ, Bipat S, van der Woude HJ. Detection of partial-thickness supraspinatus tendon tears: is a single direct MR arthrography series in ABER position as accurate as conventional MR arthrography? Skeletal Radiol 2009;38:967-975.
26. Schreinemachers SA, van der Hulst VP, Jaap Willems W, Bipat S, van der Woude HJ. Is a single direct MR arthrography series in ABER position as accurate in detecting anteroinferior labroligamentous lesions as conventional MR arthography? Skeletal Radiol 2009;38:675-683.
27. Song HT, Huh YM, Kim S, Lee SA, Kim SJ, Shin KH, et al. anterior-inferior labral lesions of recurrent shoulder dislocation evaluated by MR arthrography in an adduction internal rotation (ADIR) position. J Magn Reson Imaging 2006;23:29-35.
28. Applegate GR, Hewitt M, Snyder SJ, Watson E, Kwak S, Resnick D. Chronic labral tears: value of magnetic resonance arthrography in evaluating the glenoid labrum and labralebicipital complex. Arthroscopy 2004; 20:959-963.
29. Hayes ML, Collins MS, Morgan JA, Wenger DE, Dahm DL. Efficacy of diagnostic magnetic resonance imaging for articular cartilage lesions of the glenohumeral joint in patients with instability. Skeletal Radiol 2010;39:1199-1204.
30. Herold T, Bachthaler M, Hamer OW, Hente R, Feuerbach S, Fellner C, et al. Indirect MR arthrography of the shoulder: use of abduction and external rotation to detect full- and partial-thickness tears of the supraspinatus tendon. Radiology 2006;240:152-160.

31. Jee WH, McCauley TR, Katz LD, Matheny JM, Ruwe PA, Daigneault JP. Superior labral anterior posterior (SLAP) lesions of the glenoid labrum: reliability and accuracy of MR arthrography for diagnosis. Radiology 2001;218:127-132.
32. Oh JH, Kim JY, Choi JA, Kim WS. Effectiveness of multidetector computed tomography arthrography for the diagnosis of shoulder pathology: comparison with magnetic resonance imaging with arthroscopic correlation. J Shoulder Elbow Surg 2010;19:14-20.
33. Stetson WB, Phillips T, Deutsch A. The use of magnetic resonance arthrography to detect partial-thickness rotator cuff tears. J Bone Joint Surg Am 2005;87(2):81-88.
34. Tian CY, Cui GQ, Zheng ZZ, Ren AH. The added value of ABER position for the detection and classification of anteroinferior labroligamentous lesions in MR arthrography of the shoulder. Eur J Radiol 2013;82:651-657.
35. Toyoda H, Ito Y, Tomo H, Nakao Y, Koike T, Takaoka K. Evaluation of rotator cuff tears with magnetic resonance arthrography. Clin Orthop Relat Res 2005;439:109-115.
36. Theodoropoulos JS, Andreisek G, Harvey EJ, Wolin P. Magnetic resonance imaging and magnetic resonance arthrography of the shoulder: dependence on the level of training of the performing radiologist for diagnostic accuracy. Skeletal Radiol 2010;39:661-667.
37. Chun KA, Kim MS, Kim YJ. Comparisons of the various partial-thickness rotator cuff tears on MR arthrography and arthroscopic correlation. Korean J Radiol 2010;11:528-535.
38. Jung JY, Jee WH, Chun HJ, Ahn MI, Kim YS. Magnetic resonance arthrography including ABER view in diagnosing partial-thickness tears of the rotator cuff: accuracy, and inter- and intra-observer agreements. Acta Radiol 2010;51:194-201.
39. Duc SR, Mengiardi B, Pfirrmann CW, Jost B, Hodler J, Zanetti M. Diagnostic performance of MR arthrography after rotator cuff repair. AJR Am J Roentgenol 2006;186:237-241.
40. Guntern DV, Pfirrmann CW, Schmid MR, Zanetti M, Binkert CA, Schneeberger AG, et al. Articular cartilage lesions of the glenohumeral joint: diagnostic effectiveness of MR arthrography and prevalence in patients with subacromial impingement syndrome. Radiology 2003;226:165-170.
41. Reuss BL, Schwartzberg R, Zlatkin MB, Cooperman A, Dixon JR. Magnetic resonance imaging accuracy for the diagnosis of superior labrum anterior-inferior lesions in the community setting: eighty-three arthroscopically confirmed cases. J Shoulder Elbow Surg 2006;15:580-585.
42. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psycholog Bull 1968;70:213-220.
43. Altman DG. Statistics in medical journals: developments in the 1980s. Stat Med 1991;10:1897-1913.

Chapter | 5

Reproducibility and accuracy of instability related shoulder MRA's: Orthopaedic surgeons outperform radiologists

Susan van Grinsven, Thijs Nijenhuis, Albert van Kampen,
Peer Konings, Corné van Loon

Submitted

Abstract

Aim: To compare the diagnostic reproducibility and accuracy of radiologists with orthopaedic surgeons interpreting magnetic resonance arthrography (MRA) of instability related shoulder lesions.

Materials and methods: Two experienced musculoskeletal radiologists and two experienced orthopaedic shoulder surgeons assessed 58 MRA's for 7 instability related bony (humeral head, glenoid and greater humeral tuberosity) and soft-tissue (labroligamentous-complex, capsule and rotator cuff) shoulder lesions using a pre-specified scoring list. The first assessment was individual and the second assessment was in consensus per radiologic team or orthopaedic team. Stabilizing shoulder surgery was indicated in 45 patients. During surgery the same 7-item scoring list was used to provide the reference standard. The reproducibility between observers and the accuracy of each observer was calculated per lesion type and overall per 7 lesion types using: the Kappa coefficient; agreement percentage; sensitivity and specificity values and correct diagnosis percentage. The corresponding 95% confidence interval (CI) was calculated with the Wilson's score. Differences in agreement or correct diagnosis percentage were calculated to compare the diagnostic performance of the 2 disciplines using the McNemar test for paired proportions. A P-value of < 0.05 was considered statistically significant.

Results: The overall kappa between the individual observers ranged from fair: $k = 0.40$ (CI: 0.29-0.51) to moderate: $k = 0.49$ (CI: 0.40-0.59). Per 7 lesion types there was no significant difference ($P = 0.842$) in percentage of agreement between the individual radiologists: 78.2% (CI: 73.6-82.1) and individual orthopaedic surgeons: 79.0% (CI: 74.5-82.9). The overall sensitivity and specificity of the 4 individual observers ranged from 45.9%-76.5% and 80.1%-87.8%, respectively. The percentage of correct diagnosis of both orthopaedic surgeons: 82.0% (CI: 76.9-86.1), 75.6% (CI: 70.1-80.3) were exceedingly higher than those of their radiologic colleagues: 74.4% (CI: 68.9-79.3), 74.8% (CI: 69.3-79.7). In case of the most experienced orthopaedic surgeon this difference in accuracy was significant ($p = 0.010$ and $p = 0.018$, respectively). Furthermore, a firm non-significant trend in overall improved reproducibility and accuracy rates was detected in favour of consensus assessment compared to individual assessment.

Conclusion: Orthopaedic surgeons should rely on their own MRA interpretation if clinical diagnosis isn't confirmed by MRA and disagreement about the presence of instability related shoulder lesions exists.

Introduction

Traumatic anterior shoulder dislocation is a common injury in young (age 15-29) and active individuals. [1] The subsequent instability is often surgically treated. [2-6]

Although the diagnosis of traumatic anterior shoulder instability (TASI) is often clear based on patient history and physical examination alone [7-10], the type, location and extent of subtle bony (humeral head, glenoid and greater humeral tuberosity) and soft-tissue (labroligamentous-complex, capsule and rotator cuff) shoulder lesions may be difficult to delineate without additional imaging. [11-15] Conventional radiographs and computed tomography (CT) scans are helpful in showing bone defects, but so far, high-field magnetic resonance arthrography (MRA) with the abduction external rotation (ABER) view is the diagnostic gold standard for patients with suspected TASI. [1-7,9,11-13,15-28]

The utility of this diagnostic test depends on its accuracy (the ability of a test to discriminate between conditions of interest) and its reproducibility (the ability to achieve measurement repetition of a variable). The results of the radiologist's MRA report directly influence the orthopaedic surgeon's treatment decision (nonsurgical, arthroscopic surgery or open surgical approach) and treatment success depends on proper patient selection. [11-15]

Depending on the lesion type the sensitivity and specificity of instability related MRA assessments by experienced musculoskeletal radiologists vary from 79-100% and 85-100%, respectively [1,5,9,13-19,22,24,28-32], although much lower accuracy rates have been reported too. [3,10,23,25] The reproducibility rates of radiologists tend to be highly variable with kappa's ranging from no agreement (-0.03) to excellent agreement (0.84). [3,4,19,22,24-25,33-34] This highly variable reproducibility is further complicated by the fact that radiologists and orthopaedic surgeons do not agree about instability related lesions on MRA either, as kappa's range from -.007 to 0.86, although only a few authors studied this subject. [17,24-25,33]

In our daily clinical practice disagreement between radiologists and orthopaedic surgeons about the presence of instability related lesions on MRA frequently occurs. Consequently, the orthopaedic surgeon has to decide which MRA interpretation is likely to be the most accurate in order to plan a successful treatment strategy. The few literature available is, however, indecisive as no difference in accuracy between the 2 disciplines [17,24-25], superior [33] and inferior [10] accuracy rates of radiologists are reported.

As high reproducibility is conditional to high individual accuracy and successful treatment the purpose of the present study is to compare the diagnostic reproducibility and accuracy of experienced musculoskeletal radiologists with experienced orthopaedic shoulder surgeons interpreting high-field MRA of patients with TASI. Furthermore, we evaluated the influence of consensus assessment compared to individual assessment. Our hypotheses are that the diagnostic performance of the radiologists is superior to those of the orthopaedic surgeons and that consensus assessment will outperform individual assessment.

Materials and methods

The Rijnstate hospital is a large teaching hospital with an adherence of 425.000 inhabitants and 750 beds. All consecutive patients with traumatic anterior shoulder instability who were referred for an 1.5 T MRA in our secondary care setting from 2007 to 2011 after reduced TASI were considered for enrolment in our prospective diagnostic study. An orthopaedic surgeon and a physical therapist, both with vast experience in shoulder pathology, decided whether referral was appropriate. Exclusion criteria for patient selection were posterior, superior, multidirectional or non-traumatic instability, previous surgery of the investigated shoulder, record of a shoulder procedure between MRI and stabilizing shoulder surgery (if indicated), skeletal immaturity (age $\leq$ 16), epilepsy or a cognitive disorder. In order to perform a prospective diagnostic performance study the study protocol was designed before data collection was started and the radiologic and orthopaedic department were instructed accordingly. The regional Medical Ethical Committee informed us that in accordance with the ethical standards on human experimentation and with the Helsinki Declaration of 1964 (revised in 2004) no approval of the Ethical Committee was required and the need for informed consent was waivered considering the study type (28 June 2007: CMO 2007/141). Ultimately 58 MRA's were included.

MRA (index test)

Shoulder arthrography and MRA imaging were performed in two locations of the Rijnstate Hospital. At the first location MR imaging was performed with an 1.5 T Magnetom Avanto Tim [32x8] and a small extremity coil (Siemens AG, Munich, Germany). At the second location an 1.5 T ASC Intera Gyroscan system and a Synergy flex-M surface shoulder coil (Philips, Best, the Netherlands) was used, which was replaced during the study period by an 1.5 T Achieva system and a Synergy flex-M surface coil (Philips, Best, the Netherlands). Fluoroscopically guided and through an anterior approach a 21 gauge needle was inserted into the superior medial quadrant of the humeral head. Correct intra-articular position was confirmed with an iodinated contrast agent (2-3 cc Xenitix 300, Guerbet BV, Gorinchem, the Netherlands) after which 14-16 cc diluted gadolinium complex (Artirem, Guerbet BV, Gorinchem, the Netherlands) was injected into the glenohumeral joint. Arthrography was performed less than 30 minutes before MR imaging to obtain optimum imaging quality. Patients were placed in supine position with the arm slightly abducted and externally rotated. In case of ABER view position patients were asked to place the hand of the affected extremity posterior to the contralateral aspect of the head or neck with the elbow flexed (table 1).

Table 1: MR imaging (1.5 Tesla) protocol

Sequence	***Gradient echo/ T1-weighted (T1W/FFE/3D)***	***Turbo spin echo/ T2-weighted (T2W/TSE)***	***Proton density weighted/ Spin echo (PDW/SE)***	***Turbo spin echo/ T1-weighted (T1W/TSE)***	***Turbo spin echo/ T1-weighted (T1W/TSE)***
Orientation	Oblique coronal	Oblique coronal	Sagittal	Axial	ABER position[a]
Fat suppressed	Yes	Yes	No	Yes	Yes
Time to repeat (ms)	23	3269	1800	475	475
Time to echo (ms)	9.5	70	25	18	18
Flip angle (°)	20	90	90	90	90
Slice thickness (mm)	4	3.5	3.5	3.5	3.5
Slice gap (mm)	---	0.35	0.35	0.35	0.35
NEX[b]	3	4	2	3	3
Duration (min)	4.10	3.25	4.51	5.28	5.28
FOV[c] (mm)	170	180	130	180	180
Matrix size	304 x 228	256 x 205	288 x 202	304 x 212	256 x 205
Reconstruction matrix	512	512	512	512	512

[a]: *the non-optional abduction external rotation view improves visualization of rotator cuff partial-thickness tears and the anterior labroligamentous complex*

[b]: *number of excitations*

[c]: *field of view*

Reference standard

All MRA images were initially interpreted by a random radiologist of the 'musculoskeletal group' at our hospital. If the quality of the images was sufficient the report was send to the department of orthopaedics to confirm the clinical diagnosis and to plan treatment. Out of 58 MRA's 45 arthroscopic shoulder stabilizations were performed by 2 experienced orthopaedic shoulder surgeons. MRA reports and images were accessible during surgery. Arthroscopy was performed with regional anaesthesia and an (additional) interscalene brachial plexus block. The patient was placed in either the lateral decubitus position or the beach chair position with the arm slightly abducted and externally rotated under gentle longitudinal traction. Before incision the instability severity and instability direction were tested. During surgery a standard 4-mm 30-degree arthroscope was inserted using the classic posterior approach. Other instruments were inserted through two anterior portals. The shoulder was systematically inspected for the presence of seven instability related lesions: greater tuberosity fracture, cuff lesion (partial tear / full thickness tear), humeral head lesion (Hill-Sachs), anterior inferior glenoid lesion (bony Bankart), anterior inferior labrum lesion (classic Bankart), superior labrum anterior-to-posterior lesion (SLAP II-IV)

and joint capsule lesion (capsule redundancy / anterior stripping or tear) (figure 1-6). [6,35-40] The standard surgical scoring list served as the reference standard in our current study.

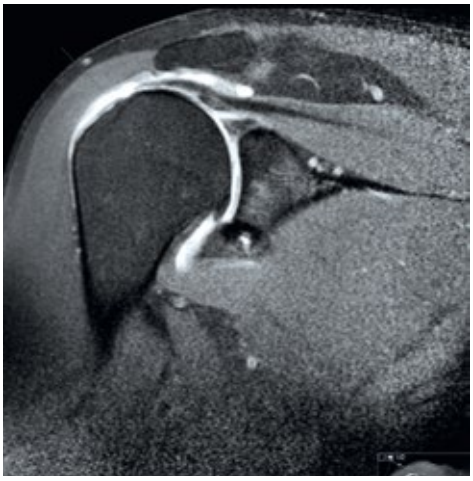

Figure 1: Arrow: Full thickness tear of the musculus Supraspinatus (cuff lesion)

Image evaluation

Four observers (2 radiologists and 2 orthopaedic surgeons) evaluated the 58 MRA images twice in a randomly ordered fashion. In the first series all observers (R1, R2, O1, O2) assessed the MRA images individually to evaluate the reproducibility between observers and the accuracy of each observer. In the second series the MRA images were assessed in consensus by team radiologists (TR) and team orthopaedic surgeons (TO) to assess the influence of consensus assessment versus individual assessment on reproducibility and accuracy.

TR consisted of the two most experienced and dedicated musculoskeletal radiologists with 16 years and 5 years of experience, respectively. Both radiologists are trained in musculoskeletal imaging and evaluate approximately 125 MRA's per radiologist per year. TO consisted of two experienced orthopaedic shoulder surgeons with 8 years and 6 years of experience in shoulder surgery, respectively. The orthopaedic surgeons of TO performed, respectively, 700 and 600 diagnostic or therapeutic shoulder arthroscopies in their career. Between the assessment of the first (individual) and the second (consensus) series a minimum interval of two months was applied to avoid MRA recognition. Each series was divided into two or three meetings to avoid fatigue. All instability related lesions were assessed according to the same seven-item scoring list as was used during surgery. The spectrum of the assessed lesions ranged from absence of pathology to the presence of multiple lesions. All series took place under comparable circumstances (e.g. room, time of

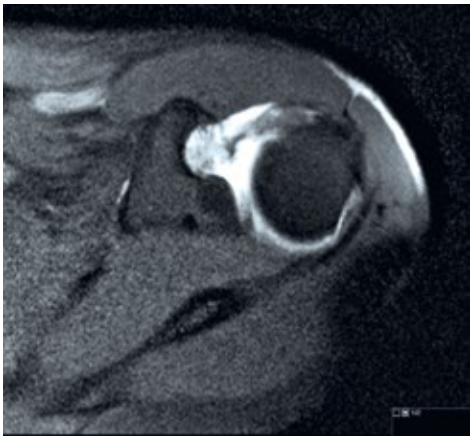

Figure 2: Arrow: Posterior-lateral bone impression fracture of the humeral head (Hill-Sachs lesion)

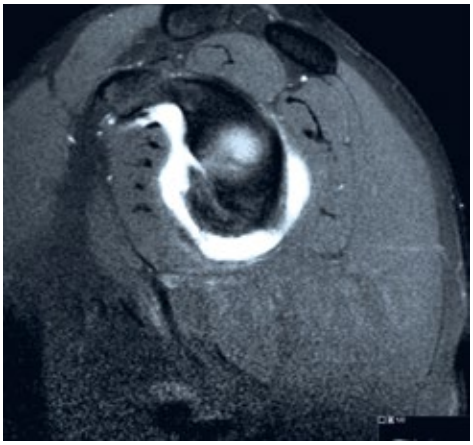

Figure 3: Arrow: Fracture of the anterior-inferior rim of the glenoid (bony Bankart)

Figure 4: Arrow: Avulsion of the anterior-inferior labrum from the rim of the glenoid (classic Bankart)

day, monitor hardware and software) and all observers were blinded for arthroscopic findings and MRA findings of other observers. All observers were, however, provided with a list of all lesion types definitions to create uniformity.

Figure 5: Arrow: Tear in the superior labrum, continues from anterior-to-superior, not depicted (SLAP lesion)

Statistical analysis

To facilitate calculations the 7 assessed lesion types were dichotomized to construct 2x2 tables. MRA reproducibility between the individual observers and between the teams was calculated using kappa coefficients [41] and the absolute percentage of agreement per different lesion type and overall per 7 lesion types. The kappa coefficient measures the degree of agreement beyond expected by chance alone. Kappa values were interpreted as: Kappa < 0, no agreement; Kappa = 0.0-0.20, poor agreement; Kappa = 0.21-0.40, fair agreement; Kappa = 0.41-0.60, moderate agreement; Kappa = 0.61-0.80, good agreement; Kappa = 0.81-1.00, excellent agreement. MRA accuracy of the individual observers and of the teams was expressed in sensitivity, specificity and percentage of correct diagnosis values as confirmed by arthroscopy per different lesion type and overall per 7 lesion types. The corresponding 95% confidence intervals (CI) were calculated using the Wilson score. Differences in absolute percentage of agreement and percentage of correct diagnosis were tested for significance using the McNemar test for paired proportions per lesion type and overall per 7 lesion types. [42] A p-value of < 0.05 was considered statistically significant. No adjustment for multiple testing was made. Data management and statistical analysis were performed using SPSS software version 18.0 (SPSS Inc., Chicago, IL, USA). Post hoc sample size calculations indicated that 40 MRA's were required to statistically detect with a power of 80% and a type 1 error of 5%, a 15% difference between the observers when the proportion of discordant pairs was conservatively assumed to be 0.65 and the method of analysis was the McNemar test of equality of paired proportions. [2]

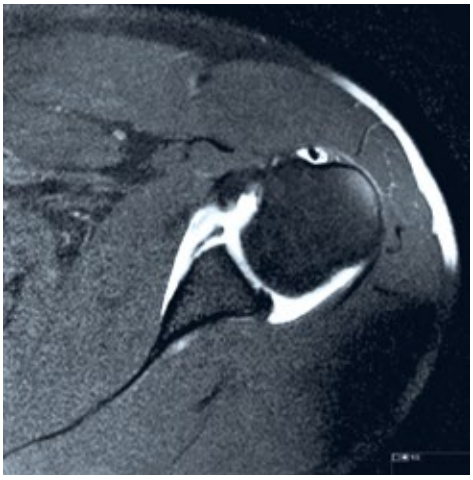

Figure 6: Arrow: Stripping of the anterior capsulo-ligamentous complex from the anterior rim of the glenoid (joint capsule lesion)

Results

Fifty-eight patients (39 men and 19 women) were enrolled in this prospective study. The mean age at the time of MRA imaging was 29.6 years (range 17-57). All MRA sequences were performed in 48 out of 58 patients (83%). The ABER position sequence was missing in 7 patients (12%) and in 3 patients (5%) one of the following sequences (coronal 2x, sagittal 1x) was missing. 2 MRA's were affected by suboptimal administration of contrast agent and 7 ABER sequences were affected by motion artefacts or malposition of the affected arm in this position. In 50 patients surgery was indicated, 5 patients refused surgery, resulting in 45 surgery confirmed MRA's (31 men and 14 women). The mean time from MRA to surgery was 4.9 months (range 1-24). All arthroscopic results and all data of the individual radiologists, individual orthopaedic surgeons and teams are complete (table 2).

MRA Reproducibility between the individual radiologists and orthopaedic surgeons (table 3 and 4)

The overall kappa between the individual observers ranged from fair: k = 0.40 (CI: 0.29-0.51) to moderate: k = 0.49 (CI: 0.40-0.59). Per 7 lesion types the absolute percentage of agreement between the individual radiologists: 78.2% (CI: 73.6-82.1) was slightly lower than between the individual orthopaedic surgeons: 79.0% (CI: 74.5-82.9). This difference was, however, not statistically different: p = 0.842. Off all seven lesion types the radiologists agreed only significant more about the presence of joint capsule lesions on MRA: R1R2: 72.5% (CI: 59.1-82.9) versus O1O2: 35.3% (CI: 23.6-49.0), p = 0.001. Moreover, the radiologists were outperformed by the orthopaedic surgeons in case of SLAP lesions: R1R2: 58.8% (CI: 45.2-71.2) versus O1O2: 90.2% (CI: 79.0-95.7), p < 0.001.

Table 2: Frequency of pathologies seen on MRA or during surgery

Items standard scoring list		***R1*** ***N=51 / N=38***	***R2*** ***N=51 / N=38***	***O1*** ***N=51 / N=38***	***O2*** ***N=51 / N=38***
Cuff lesion:					
	Not present	41 / 31	35 / 27	43 / 31	41 / 30
	Partial tear	10 / 7	9 / 6	8 / 7	9 / 7
	Full thickness tear	0 / 0	7 / 5	0 / 0	1 / 1
Humeral head lesion:					
	Not present	14 / 9	10 / 6	11 / 7	16 / 11
	Degeneration	10 / 7	6 / 6	5 / 4	0 / 0
	Hill-Sachs	27 / 22	35 / 26	35 / 27	35 / 27
Anterior inferior glenoid lesion:					
	Not present	45 / 35	42 / 33	46 / 35	49 / 37
	Degeneration	4 / 1	6 / 3	3 / 2	0 / 0
	Defect/Bony Bankart	2 / 2	3 / 2	2 / 1	2 / 1
Greater humeral tuberosity fractures:					
	Not present	51 / 38	51 / 38	50 / 37	50 / 37
	Present	0 / 0	0 / 0	1 / 1	1 / 1
Anterior inferior labrum lesion:					
	Not present	30 / 22	12 / 9	9 / 6	24 / 16
	Degeneration/Fraying	6 / 2	0 / 0	0 / 0	0 / 0
	Classic Bankart	15 / 14	39 / 29	42 / 32	27 / 22
SLAP lesion:					
	Not present	45 / 32	30 / 24	50 / 37	47 / 35
	Present	6 / 6	21 / 14	1 / 1	4 / 3
Joint capsule lesion:					
	Not present	34 / 28	38 / 30	19 / 14	46 / 34
	Capsule redundancy	14 / 10	2 / 1	4 / 3	2 / 2
	Anterior stripping/Tear	3 / 0	11 / 7	28 / 21	3 / 2

Items standard scoring list		*TR N=58 / N=45*	*TO N=58 / N=45*	*Surgical frequency*
Cuff lesion:				
	Not present	48 / 38	53 / 42	40
	Partial tear	7 / 4	5 / 3	4
	Full thickness tear	3 / 3	0 / 0	1
Humeral head lesion:				
	Not present	11 / 8	20 / 14	17
	Degeneration	9 / 5	2 / 2	0
	Hill-Sachs	38 / 32	36 / 29	28
Anterior inferior glenoid lesion:				
	Not present	53 / 43	55 / 44	35
	Degeneration	2 / 0	2 / 1	7
	Defect/Bony Bankart	3 / 2	1 / 0	3
Greater humeral tuberosity fractures:				
	Not present	58 / 45	58 / 45	45
	Present	0 / 0	0 / 0	0
Anterior inferior labrum lesion:				
	Not present	15 / 10	10 / 5	4
	Degeneration/Fraying	6 / 4	5 / 3	3
	Classic Bankart	37 / 31	43 / 37	38
SLAP lesion:				
	Not present	41 / 34	56 / 44	43
	Present	17 / 11	2 / 1	2
Joint capsule lesion:				
	Not present	47 / 37	49 / 42	14
	Capsule redundancy	8 / 6	0 / 0	14
	Anterior stripping/Tear	3 / 2	9 / 3	17

MRA: magnetic resonance arthrography, SLAP: superior labral anterior-to-posterior lesion, ABER: abduction external rotation view, R1: radiologist 1, R2: radiologist 2, O1: orthopaedic surgeon 1, O2: orthopaedic surgeon 2, TR: team radiologists, TO: team orthopaedic surgeons, N=58: all MRA's, N=51: all MRA's with ABER, N=45: surgically confirmed MRA's, N=38: surgically confirmed MRA's with ABER

Table 3: Reproducibility of MRA diagnosis between the individual observers (N=51, 357 pair-wise ratings) and teams (N=58, 406 pair-wise ratings)

	R1/O1	***R1/O2***	***R2/O1***	***R2/O2***
Cuff lesion≈	0.46 / 84.3	0.38 / 80.4	0.37 / 76.5	0.29 / 72.5
Humeral head lesion≈	0.44 / 72.5	0.36 / 68.6	0.64 / 84.3	0.55 / 80.4
Anterior inferior glenoid lesion≈	0.48 / 96.1	0.48 / 96.1	0.79 / 98.0	0.79 / 98.0
Greater humeral tuberosity fractures≈	NA / 98.0	NA / 98.0	NA / 98.0	NA / 98.0
Anterior inferior labrum lesion≈	0.16 / 47.1	0.12 / 84.3	0.23 / 74.5	0.11 / 56.7
SLAP lesion≈	0.26 / 90.2	0.31 / 64.7	-0.04 / 56.7	0.22 / 66.7
Joint capsule lesion≈	0.24 / 58.8	-0.07 / 60.8	0.06 / 47.1	-0.04 / 68.6
Overall absolute % of agreement (95% *CI*)	**78.2 (73.6-82.1)**	**79.0 (74.5-82.9)**	**76.5 (71.8-80.6)**	**77.3 (72.7-81.4)**
Overall kappa (95% *CI*)	**0.47 (0.37-0.56)**	**0.40 (0.29-0.51)**	**0.48 (0.39-0.58)**	**0.46 (0.37-0.56)**

	R1/R2	***O1/O2***	***TR/TO***
Cuff lesion≈	0.70 / 88.2	0.46 / 84.3	0.32 / 84.5
Humeral head lesion≈	0.60 / 80.4	0.55 / 80.4	0.55 / 79.3
Anterior inferior glenoid lesion≈	0.37 / 94.1	1.00 / 100.0	-0.03 / 93.1
Greater humeral tuberosity fractures≈	NA / 100.0	-0.02 / 96.1	NA / 100.0
Anterior inferior labrum lesion≈	0.23 / 52.9	0.31 / 66.7	0.28 / 69.0
SLAP lesion≈	0.05 / 58.8	-0.03 / 90.2	0.05 / 70.7
Joint capsule lesion≈	0.34 / 72.5	-0.07 / 35.3	0.28 / 79.3
Overall absolute % of agreement (95% *CI*)	**78.2 (73.6-82.1)**	**79.0 (74.5-82.9)**	**82.3 (78.3-85.7)**
Overall kappa (95% *CI*)	**0.48 (0.38-0.57)**	**0.49 (0.40-0.59)**	**0.54 (0.45-0.63)**

NA: not applicable (k-coefficient could not be determined due to empty cells in 2x2 table), MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, R1: radiologist 1, R2: radiologist 2, O1: orthopaedic surgeon 1, O2: orthopaedic surgeon 2, TR: team radiologists, TO: team orthopaedic surgeons, ≈: kappa-coefficient / absolute percentage of agreement, CI: (confidence interval)

MRA Accuracy of the individual radiologists and orthopaedic surgeons (table 5 and 6)

The overall sensitivity and specificity of the individual observers ranged from 45.9%-76.5% and 80.1%-87.8%, respectively. Per 7 lesion types both orthopaedic surgeons: 82.0% (CI: 76.9-86.1), 75.6% (CI: 70.1-80.3) were more accurate than the radiologists: 74.4% (CI: 68.9-79.3), 74.8% (CI: 69.3-79.7). In case of the most experienced orthopaedic surgeon O1 this difference in percentage of correct diagnosis was significant (p = 0.010 and p = 0.018, respectively). Per lesion type the orthopaedic surgeons were especially more

Table 4: Influence of consensus assessment on the reproducibility of MRA diagnosis (N=51, 357 pair-wise ratings)

	R1R2 / O1O2	***TRTO / R1R2***	***TRTO / O1O2***
Cuff lesion	0.754	0.791	1.000
Humeral head lesion	1.000	1.000	1.000
Anterior inferior glenoid lesion	NA	1.000	NA
Greater humeral tuberosity fractures	NA	NA	NA
Anterior inferior labrum lesion	0.210	[c] 0.031	0.648
SLAP lesion	[a] < 0.001	0.267	[d] 0.007
Joint capsule lesion	[b] 0.001	0.648	[e] < 0.001
Overall difference in absolute % of agreement	**0.842**	**0.115**	**0.266**
Which observers scores best?	**O1O2**	**TRTO**	**TRTO**

P-values per lesion type and overall per 7 lesion types (differences in % of agreement between the paired individual radiologists and paired orthopaedic surgeons or between the paired teams and their consisting paired individual observers are tested using the McNemar test for paired proportions), NA: not applicable (P-values could not be determined due to empty cells in 2x2 table), MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, R1: radiologist 1, R2: radiologist 2, O1: orthopaedic surgeon 1, O2: orthopaedic surgeon 2, TR: team radiologists, TO: team orthopaedic surgeons, [a]p < 0.01 versus R1R2, [b]p < 0.01 versus O1O2, [c]p < 0.05 versus R1R2, [d]p < 0.01 versus TRTO, [e]p < 0.01 versus O1O2

accurate in case of anterior inferior labrum lesions: O1: 86.8% (CI: 72.7-94.3) versus R1: 50.0% (CI: 34.9-65.2), p = 0.001 and SLAP lesions: O1: 92.1% (CI: 79.2-97.3) versus R2 63.2% (CI: 47.3-76.6), p = 0.007 and O2: 86.8% (CI: 72.7-94.3) versus R2: 63.2% (CI: 47.3-76.6), p = 0.012.

The influence of the consensus assessments of the teams on MRA reproducibility (table 3 and 4)

Per 7 lesion types the kappa and absolute percentage of agreement between the teams: k = 0.54 (CI: 0.45-0.63), 82.3% (CI: 78.3-85.7) was exceedingly higher than the overall agreement between the individual radiologists: k = 0.48 (CI: 0.38-0.57), 78.2% (CI: 73.6-82.1) or individual orthopaedic surgeons: k = 0.49 (CI: 0.40-0.59), 79.0% (CI: 74.5-82.9). In fact the overall agreement between the teams (k = 0.54, 82.3%) was higher than between all possible combinations of individual observers (range: k = 0.40-0.49, 76.5%-79.0%). Although we established a firm trend of superior overall reproducibility through consensus assessment significant superior agreement in favour of the teams was only found in case of anterior inferior labrum lesions: TRTO: 69.0% (CI: 56.2-79.4) versus

Table 5: Accuracy of MRA diagnosis (%) of the individual radiologists and orthopedic surgeons, per lesion type and overall per 7 lesion types

			R1			***R2***
	Sens	Spec	Acc (95% *CI*)	Sens	Spec	Acc (95% *CI*)
Cuff lesion	80.0	90.9	89.5 (75.9-95.8)	80.0	78.8	78.9 (63.7-88.9)
Humeral head lesion	68.2	56.3	63.2 (47.3-76.6)	77.3	43.8	63.2 (47.3-76.6)
Anterior inferior glenoid lesion	0	94.6	92.1 (79.2-97.3)	100	97.3	97.4 (86.5-99.5)
Greater humeral tuberosity fractures	NA	100	100 (90.8-100)	NA	100	100 (90.8-100)
Anterior inferior labrum lesion	41.9	85.7	50.0 (34.9-65.2)	80.6	42.9	73.7 (58.0-85.0)
SLAP lesion	50.0	86.1	84.2 (69.6-92.6)	50.0	63.9	63.2 (47.3-76.6)
Joint capsule lesion	25.0	71.4	42.1 (27.9-57.8)	25.0	85.7	47.4 (32.5-62.7)
Overall Sens, Spec and Acc (N=38, 266 pair-wise ratings)	**45.9**	**87.8**	**74.4 (68.9-79.3)**	**63.5**	**80.1**	**74.8 (69.3-79.7)**
			O1			***O2***
	Sens	Spec	Acc (95% *CI*)	Sens	Spec	Acc (95% *CI*)
Cuff lesion	60.0	87.9	84.2 (69.6-92.6)	60.0	84.8	81.6 (66.6-90.8)
Humeral head lesion	77.3	37.5	60.5 (44.7-74.4)	72.7	31.3	55.3 (39.7-69.9)
Anterior inferior glenoid lesion	0	97.3	94.7 (82.7-98.5)	0	97.3	94.7 (82.7-98.5)
Greater humeral tuberosity fractures	NA	97.4	97.4 (86.5-99.5)	NA	97.4	97.4 (86.5-99.5)
Anterior inferior labrum lesion	93.5	57.1	86.8 (72.7-94.3)	67.7	85.7	71.1 (55.2-83.0)
SLAP lesion	0	97.2	92.1 (79.2-97.3)	0	91.7	86.8 (72.7-94.3)
Joint capsule lesion	66.7	42.9	57.9 (42.2-72.2)	12.5	92.9	42.1 (27.9-57.8)
Overall Sens, Spec and Acc (N=38, 266 pair-wise ratings)	**76.5**	**84.5**	**82.0 (76.9-86.1)**	**50.6**	**87.3**	**75.6 (70.1-80.3)**
			TR			***TO***
	Sens	Spec	Acc (95% *CI*)	Sens	Spec	Acc (95% *CI*)
Cuff lesion	80.0	92.5	91.1 (79.3-96.5)	40.0	97.5	91.1 (79.3-96.5)
Humeral head lesion	78.6	41.2	64.4 (49.8-76.8)	71.4	47.1	62.2 (47.6-74.9)
Anterior inferior glenoid lesion	0	95.2	88.9 (76.5-95.2)	0	100	93.3 (82.1-97.7)
Greater humeral tuberosity fractures	NA	100	100 (92.1-100)	NA	100	100 (92.1-100)
Anterior inferior labrum lesion	73.7	57.1	71.1 (56.6-82.3)	89.5	57.1	84.4 (71.2-92.3)
SLAP lesion	100	79.1	80.0 (66.2-89.1)	0	97.7	93.3 (82.1-97.7)
Joint capsule lesion	16.1	78.6	35.6 (23.2-50.2)	6.5	92.9	33.3 (21.4-47.9)
Overall Sens, Spec and Acc (N=45, 315 pair-wise ratings)	**57.0**	**85.6**	**75.9 (70.9-80.3)**	**54.2**	**92.8**	**79.7 (74.9-83.8)**

NA: not applicable (accuracy rates could not be determined due to empty cells in 2x2 table), MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, R1: radiologist 1, R2: radiologist 2, O1: orthopaedic surgeon 1, O2: orthopaedic surgeon 2, TR: team radiologists, TO: team orthopaedic surgeons, Sens: sensitivity, Spec: specificity, Acc: percentage of correct diagnosis values calculated as: (true-positive + true-negative) / total patients in 2x2 table, CI: (confidence interval)

Table 6: Difference in percentage correctly diagnosed between the individual observers (N=38, 266 pair-wise ratings)

	R1 / R2	***R1 / O1***	***R1 / O2***
Cuff lesion	0.125	0.687	0.453
Humeral head lesion	1.000	1.000	0.549
Anterior inferior glenoid lesion	0.500	1.000	1.000
Greater humeral tuberosity fractures	NA	NA	NA
Anterior inferior labrum lesion	[a] 0.035	[b] 0.001	0.057
SLAP lesion	0.057	0.375	1.000
Joint capsule lesion	0.727	0.210	1.000
Overall difference in % correctly diagnosed	**1.000**	[c] **0.010**	**0.784**
Which observer scores best?	**R2**	**O1**	**O2**
	R2 / O1	***R2 / O2***	***O1 / O2***
Cuff lesion	0.727	1.000	1.000
Humeral head lesion	1.000	0.453	0.727
Anterior inferior glenoid lesion	1.000	1.000	1.000
Greater humeral tuberosity fractures	NA	NA	1.000
Anterior inferior labrum lesion	0.180	1.000	0.146
SLAP lesion	[d] 0.007	[f] 0.012	0.625
Joint capsule lesion	0.503	0.754	0.307
Overall difference in % correctly diagnosed	[e] **0.018**	**0.894**	[g] **0.033**
Which observer scores best?	**O1**	**O2**	**O1**

P-values per lesion type and overall per 7 lesion types (tested using the McNemar test for paired proportions), NA: not applicable (P-values could not be determined due to empty cells in 2x2 table), SLAP: superior labrum anterior-to-posterior lesion, R1: radiologist 1, R2: radiologist 2, O1: orthopaedic surgeon 1, O2: orthopaedic surgeon 2, [a]p < 0.05 versus R1, [b]p < 0.01 versus R1, [c]p < 0.05 versus R1, [d]p < 0.01 versus R2, [e]p < 0.05 versus R2, [f]p < 0.05 versus R2, [g]p < 0.05 versus O2

R1R2: 52.9% (CI: 39.5-66.0), p = 0.031 and joint capsule lesions: TRTO: 79.3% (CI: 67.2-87.8) versus O1O2: 35.3% (CI: 23.6-49.0), p < 0.001.

The influence of the consensus assessments of the teams on MRA accuracy (table 5 and 7)

The overall sensitivity and specificity of the teams ranged from 54.2%-57.0% and 85.6%-92.8%, respectively. Per 7 lesion types the orthopaedic team: 79.7% (CI: 74.9-83.8) was more accurate than the radiologic team: 75.9% (CI: 70.9-80.3). This difference was, however, not statistically significant: p = 0.104. The overall accuracy of the teams was superior to the overall accuracy of the individual observers in 3 out of 4 times. Team

Table 7: Influence of consensus assessment on the accuracy of MRA diagnosis

	TR / TO	***TR / R1***	***TR / R2***
Cuff lesion	1.000	1.000	0.125
Humeral head lesion	1.000	1.000	1.000
Anterior inferior glenoid lesion	0.500	1.000	1.000
Greater humeral tuberosity fractures	NA	NA	NA
Anterior inferior labrum lesion	0.146	[a] 0.022	1.000
SLAP lesion	0.146	0.727	0.070
Joint capsule lesion	1.000	1.000	0.581
Overall difference in % correctly diagnosed	**0.104**	**0.291**	**0.349**
Which observer(s) scores best?	**TO**	**TR**	**TR**
	TO / O1	***TO / O2***	
Cuff lesion	0.687	0.453	
Humeral head lesion	1.000	0.508	
Anterior inferior glenoid lesion	1.000	1.000	
Greater humeral tuberosity fractures	NA	NA	
Anterior inferior labrum lesion	1.000	0.227	
SLAP lesion	1.000	0.625	
Joint capsule lesion	0.189	1.000	
Overall difference in % correctly diagnosed	**0.761**	[b] **0.034**	
Which observer(s) scores best?	**O1**	**TO**	

P-values per lesion type and overall per 7 lesion types (differences in % correctly diagnosed between the teams (N=45, 315 pair-wise ratings) and between the teams and their individual observers (N=38, 266 pair wise ratings) are tested using the McNemar test for paired proportions), NA: not applicable (P-values could not be determined due to empty cells in 2x2 table), MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, R1: radiologist 1, R2: radiologist 2, O1: orthopaedic surgeon 1, O2: orthopaedic surgeon 2, TR: team radiologists, TO: team orthopaedic surgeons; [a]p <0.05 versus R1, [b]p < 0.05 versus O2

radiologists: 75.9% (CI: 0.9-80.3) was more accurate than both individual radiologists: 74.4% (CI: 68.9-79.3), 74.8% (CI: 69.3-79.7) and team orthopaedic surgeons: 79.7% (CI: 74.9-83.8) significantly improved the accuracy of its weakest member O2: 75.6% (CI: 70.1-80.3), p = 0.034. Per lesion type superior significant accuracy improvement through consensus assessment was only achieved in case of anterior inferior labrum lesions: TR: 71.1% (CI: 56.6-82.3) versus R1: 50.5% (CI: 34.9-65.2), p = 0.022.

Discussion

In contradiction to our expectation the diagnostic overall reproducibility and accuracy of the radiologists was not superior to that of the orthopaedic shoulder surgeons. The orthopaedic surgeons actually agreed slightly more and were more accurate in the assessment of seven instability related shoulder lesions on MRA than the experienced musculoskeletal radiologists. In accordance to our hypothesis we detected a firm trend indicating, that assessment in consensus improves the overall diagnostic performance of the individual radiologists and orthopaedic surgeons.

In the present study the overall reproducibility between the radiologists (k = 0.48) and orthopaedic surgeons (k = 0.49) were both moderate and there was no overall significant difference in the percentage of agreement between the two disciplines (78.2% / 79.0%, p = 0.842). Although just a few other authors compared the reproducibility between radiologist and orthopaedic surgeons before [17,24-25,33] our reproducibility rates are perfectly in line with the results mentioned in the 2 studies of Schreinemachers et al.. In their study an experienced orthopaedic surgeon and 2 experienced radiologists evaluated cuff and anterior inferior labrum lesions on 92 MRA's. The reported agreement between the two disciplines was only moderate too (k = 0.48-0.56 and k = 0.44-0.56, respectively). In spite of the fact that we provided all individual observers with the same lesion definitions list it seems that the 2 disciplines differ in definition interpretation or use of terminology, as was earlier suggested by other authors. [8, 33]

Although the overall specificity rates (range: 80.1%-87.8%) of the 4 individual observers were also in line with the results in previous literature studying MRA assessments of radiologists the overall sensitivity rates of the orthopaedic surgeons (50.6%, 76.5%) and radiologists (45.9%, 63.5%) were lower than generally reported (79%-100%, 85%-100%). [1,5,9,13-19,22,24,28-32] The overall accuracy rates of the present study are, however, in perfect concurrence with a former retrospective study of ours [3] where the same 7 lesion types were assessed by 2 radiologists on 61 MRA's resulting in a sensitivity of 50.0% to 60.6% and a specificity of 66.8% to 87.9%. Overall, the most experienced orthopaedic surgeon (82.0%) was significant more accurate than both radiologists (74.4%, 74.8%) in the present study. Per lesion type the orthopaedic surgeons were especially more accurate in the assessment of anterior inferior labrum and SLAP lesions on MRA. As only a few authors compared the accuracy of radiologists with orthopaedic surgeons before it is notable that the results of the study of Pandaya et al. confirm our findings. [10] Here, the orthopaedic surgeon (sensitivity: 72%) was also more accurate than the radiologist (sensitivity: 50%) in the assessment of SLAP lesions on 51 MRA's.

An explanation for the superior overall accuracy of orthopaedic surgeons could be that they have a better trained three-dimensional orientation than the radiologists. During arthroscopic surgery they have the opportunity to visualize the anatomically complex

structures of the shoulder from various distances and angles. The radiologists are merely trained in assessing two-dimensional pictures. Furthermore, the orthopaedic surgeons have the advantage of direct personal feedback by verifying their MRA interpretations during surgery. Systematic personal feedback of the radiologist's MRA report after stabilizing surgery is not customary in the daily clinical practice of our hospital. At last, radiologists are focussed on every abnormality or defect seen on MRA while orthopaedic surgeons merely focus on instability related lesions. The superior accuracy of orthopaedic surgeons in the assessment of SLAP and anterior inferior labrum lesions could be caused by suspicion for these lesions based on clinical experience (prevalence of pathology), as suggested earlier. [10] The above mentioned discrepancies between the two disciplines might have led to different learning curves and consequently to differences in overall diagnostic accuracy.

In the present study the overall agreement between the consensus assessments of the teams (82.3%) was much higher than the overall agreement between the individual radiologists (78.2%) or orthopaedic surgeons (79.0%). In fact it was higher than those between all possible combinations of individual observers (range 76.5%-79.0%). Per 7 lesion types team radiologists (75.9%) was more accurate than both individual radiologists (74.4%, 74.8%) and team orthopaedic surgeons (79.7%) improved the overall accuracy of its weakest member (O2: 75.6%). Although the differences were not often statistically different a firm trend of superior overall diagnostic performance through consensus assessment was established. Apparently, diagnostic shortcomings of a radiologist or orthopaedic surgeon are complemented by the other team member. Unfortunately, there are no prospective diagnostic studies on this topic available to support our results. Future research should, therefore, further determine the added value of consensus assessment and its cost effectiveness.

As sufficient high reproducibility is conditional to systematic high individual accuracy the moderate overall reproducibility and slightly disappointing overall sensitivity rates of the radiologists and orthopaedic surgeons in this study could lead to diagnoses disagreement, questionable treatment decisions and ultimately to suboptimal health care. Continues improvement of instability related MRA performance is, therefore, essential and could be accomplished by various measures. First of all, continuous technical development in MRA imaging should be promoted (i.e. improvement of MRA sequence setting, contrast solution and monitor resolution) to facilitate easier MRA assessment. Furthermore, intensified personal feedback on MRA assessment and interdisciplinary collaboration session should be organized after stabilizing surgery to discuss discrepancies, to fine-tune definition interpretation agreement and to exchange knowledge. In the future, it would be useful to conduct a planned education of musculoskeletal radiologists in the form of arthroscopic shoulder surgery participation that would provide for a better three-dimensional orientation and direct personal feedback. Finally, even experienced dedicated musculoskeletal radiologists or experienced shoulder orthopaedic surgeons

should be stimulated to consult each other in case of hard to diagnose subtle instability related shoulder lesions on MRA.

A limitation of our study could be that we selected the 4 observers based on their experience level. Although we classified the available radiologists and orthopaedic surgeons to the best of our abilities the choices we made could have affected the results of our study. Secondly, the decision to perform surgery was based on clinical examination and MRA findings. Therefore, incorporation bias might have been introduced. Thirdly, although arthroscopy is the best available reference standard it is still a surgeon-dependent method. At last, 3 MRI systems were used and a total of 10 MRI sequences were not performed (7 ABER-, 1 sagittal- and 2 T1 coronal oblique sequences). This could have negatively affected the results of the current study. The circumstances were, however, equal for all observers, representative of normal clinical practice and the inability to assume the ABER position in 7 patients is in concurrence with previous literature. [5,24-25] We do not think that the additional surplus of 3 missing sequences out of 290 sequences (58 MRA's x 5 sequences = 290) have caused significant deviations in our results. A strength of the current study is, that it is the first prospective study evaluating the influence of discipline type and consensus assessment on diagnostic reproducibility and accuracy with sufficient statistical power and fully described definitions of the seven assessed lesions to create uniformity among observers.

Conclusion

In conclusion we state that the results of our study indicate that orthopaedic surgeons should rely on their own MRA interpretation if the clinical diagnosis is not confirmed by MRA and disagreement between the radiologist and orthopaedic surgeon exists about the presence of instability related shoulder lesions on MRA. Continuous improvement of diagnostic performance through technical development of MRA imaging, intensified personal feedback on MRA assessment, interdisciplinary collaboration and assessment in consensus in case of difficult to diagnose lesions should however be advocated. Higher diagnostic reproducibility will consequently lead to higher individual diagnostic accuracy, increased agreement about treatment strategies and ultimately to better quality of healthcare.

References

1. Khedr SA, Kassem HM, Azab MA. Comparative study of direct MR arthrography and CT arthrography with arthroscopic correlation in preoperative evaluation of anterior shoulder instability. Egypt J Radiol Nucl Med 2013;44:817–827.
2. Acid S, Le Corroller T, Aswad R, Pauly V, Champsaur P. Preoperative imaging of anterior shoulder instability: diagnostic effectiveness of MDCT arthrography and comparison with MR arthrography and arthroscopy. AJR Am J Roentgenol 2012;198(3):661-667.
3. van Grinsven S, Kesselring FO, van Wassenaer-van Hall HN, Lindeboom R, Lucas C, van Loon CJ. MR arthrography of traumatic anterior shoulder lesions showed modest reproducibility and accuracy when evaluated under clinical circumstances. Arch Orthop Trauma Surg 2007;127(1):11-17.
4. Simao MN, Nogueira-Barbosa MH, Muglia VF, Barbieri CH. Anterior shoulder instability: correlation between magnetic resonance arthrography, ultrasound arthrography and intraoperative findings. Ultrasound Med Biol 2012;38(4):551-560.
5. Smith TO, Drew BT, Toms AP. A meta-analysis of the diagnostic test accuracy of MRA and MRI for the detection of glenoid labral injury. Arch Orthop Trauma Surg 2012;132(7):905-919.
6. van der Woude HJ, Vanhoenacker FM. MR Arthrography in Glenohumeral Instability. JBR–BTR 2007;90:377-383.
7. Bak K, Wiesler ER, Poehling GG. Consensus statement on shoulder instability.Arthroscopy 2010;26(2):249-255.
8. van Kampen DA, van den Berg T, van der Woude HJ, Castelein RM, Terwee CB, Willems WJ. Diagnostic value of patient characteristics, history, and six clinical tests for traumatic anterior shoulder instability. J Shoulder Elbow Surg 2013;22:1310-1319.
9. Oh JH, Kim Y, Choi JA, Kim WS. Effectiveness of multidetector computed tomography arthrography for the diagnosis of shoulder pathology: comparison with magnetic resonance imaging with arthroscopic correlation. J Shoulder Elbow Surg 2010;19(1):14-20.
10. Pandya NK, Colton A, Webner D, Sennett B, Huffman GR. Physical examination and magnetic resonance imaging in the diagnosis of superior labrum anterior-posterior lesions of the shoulder: a sensitivity analysis. Arthroscopy 2008;24(3):311-317.
11. Amin MF, Youssef AO. The diagnostic value of magnetic resonance arthrography of the shoulder in detection and grading of SLAP lesions: Comparison with arthroscopic findings. Eur J Radiol 2012;81(9):2343-2347.
12. Bencardino J, Gyftopoulos S, Palmer W. Imaging in Anterior Glenohumeral Instability. Radiology 2013;269(2): 323-337.
13. Genovese E, Spano E, Castagna A, Leonardi A, Angeretti MG, Callegari L, Fugazzola C. MR-arthrography in superior instability of the shoulder: correlation with arthroscopy. Radiol med 2013;118:1022–1033.
14. Iqbal HJ, Rani S, Mahmood A, Brownson P, Aniq H. Diagnostic value of MR arthrogram in SLAP lesions of the shoulder. Surgeon 2010;8(6):303-309.
15. de Jesus JO, Parker L, Frangos AJ, Nazarian LN. Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. AJR Am J Roentgenol 2009;192(6):1701-1707.

16. Tian CY, Cui GQ, Zheng ZZ, Ren AH. The added value of ABER position for the detection and classification of anteroinferior labroligamentous lesions in MR arthrography of the shoulder. Eur J Radiol 2013;82(4):651-657.
17. Dinter DJ, Martetschläger F, Büsing KA, Schönberg SO, Scharf HP, Lehmann LJ. Shoulder Injuries in Overhead Athletes: Utility of MR Arthrography. Sportverl Sportschad 2008;22:146–152.
18. Drescher R, von Rothenburg T, Ludwig J, Köster O, Schmid G. Direct MR-arthrography of the shoulder with maximum capsular distension for surgical planning. Fortschr Röntgenstr 2004;176(10):1469–1474.
19. Holzapfel K, Waldt S, Bruegel M, Paul J, Heinrich P, Imhoff AB, Rummeny EJ, Woertler K. Inter- and intraobserver variability of MR arthrography in the detection and classification of superior labral anterior posterior (SLAP) lesions: evaluation in 78 cases with arthroscopic correlation. Eur Radiol 2010;20(3):666-673.
20. Ito Y, Sakai T, Tomo H, Nakao Y, Inui K, Koike T, Nakatsuchi T, Takaoka K. Computerized assessment of Bankart lesions under tension with magnetic resonance arthrography. J Shoulder Elbow Surg 2005;14(3):247-251.
21. Jonas SC, Walton MJ, Sarangi PP. Is MRA an unnecessary expense in the management of a clinically unstable shoulder? Acta Orthop 2012;83(2):1-4.
22. Jung JY, Jee WH, Chun HJ, Ahn MI, Kim YS. Magnetic resonance arthrography including ABER view in diagnosing partial-thickness tears of the rotator cuff: accuracy, and inter- and intra-observer agreements. Acta Radiol 2010;51(2):194-201.
23. Reuss BL, Schwartzberg R, Zlatkin MB, Cooperman A, Dixon JR. Magnetic resonance imaging accuracy for the diagnosis of superior labrum anterior-posterior lesions in the community setting: eighty-three arthroscopically confirmed cases. J Shoulder Elbow Surg 2006;15(5):580-585.
24. Schreinemachers SA, van der Hulst VP, Willems WJ, Bipat S, van der Woude HJ. Is a single direct MR arthrography series in ABER position as accurate in detecting anteroinferior labroligamentous lesions as conventional MR arthography? Skeletal Radiol 2009;38(7):675-683.
25. Schreinemachers SA, van der Hulst VP, Willems WJ, Bipat S, van der Woude HJ. Detection of partial-thickness supraspinatus tendon tears: is a single direct MR arthrography series in ABER position as accurate as conventional MR arthrography? Skeletal Radiol 2009;38(10):967-75.
26. Song HT, Huh YM, Kim S, Lee SA, Kim SJ, Shin KH, Suh JS. Anterior inferior labral lesions of recurrent shoulder dislocation evaluated by MR arthrography in an adduction internal rotation (ADIR) position. J Magn Reson Imaging 2006;23(1):29-35.
27. Stetson WB, Phillips T, Deutsch A. The use of magnetic resonance arthrography to detect partial-thickness rotator cuff tears. J Bone Joint Surg Am 2005;87:81-88.
28. Theodoropoulos JS, Andreisek G, Harvey EJ, Wolin P. Magnetic resonance imaging and magnetic resonance arthrography of the shoulder: dependence on the level of training of the performing radiologist for diagnostic accuracy. Skeletal Radiol 2010;39(7):661-667.
29. Toyoda H, Ito Y, Tomo H, Nakao Y, Koike T, Takaoka K. Evaluation of rotator cuff tears with magnetic resonance arthrography. Clin Orthop Relat Res 2005;439:109-115.
30. Applegate GR, Hewitt M, Snyder SJ, Watson E, Kwak S, Resnick D. Chronic labral tears: value of magnetic resonance arthrography in evaluating the glenoid labrum and labral-bicipital complex. Arthroscopy 2004;20(9):959-963.

31. Hayes ML, Collins MS, Morgan JA, Wenger DE, Dahm DL. Efficacy of diagnostic magnetic resonance imaging for articular cartilage lesions of the glenohumeral joint in patients with instability. Skeletal Radiol 2010;39(12):1199-1204.
32. Jana M, Srivastava DN, Sharma R, Gamanagatti S, Nag HL, Mittal R, Upadhyay AD. Magnetic resonance arthrography for assessing severity of glenohumeral labroligamentous lesions. J Orthop Surg 2012;20(2):230-235.
33. Halma JJ, Eshuis R, Krebbers YM, Weits T, de Gast A. Interdisciplinary inter-observer agreement and accuracy of MR imaging of the shoulder with arthroscopic correlation. Arch Orthop Trauma Surg 2012; 132(3):311-320.
34. van der Veen HC, Collins JP, Rijk PC. Value of magnetic resonance arthrography in post-traumatic anterior shoulder instability prior to arthroscopy: a prospective evaluation of MRA versus arthroscopy. Arch Orthop Trauma Surg 2012;132(3):371-375.
35. Blum A, Coudane H, Molé D. Gleno-humeral instabilities. Eur Radiol 2000;10:63-82.
36. Matsen FA, Rockwood CA. Glenohumeral instability. The Shoulder. Matsen FA, Rockwood CA, editors. WB Saunder Company, Philadelphia 1998; 611-745.
37. Sailer J, Imhof H. Schultergelenkinstabilität. Radiologe 2004;44:578-590.
38. Sanders TG, Morisson WB, Miller MD. Imaging Techniques for the Evaluation of Glenohumeral instability. Am J Sports Med 2000;28:414-434.
39. Snyder SJ, Karzel RP, Del Pizzo W, Ferkel RD, Friedman MJ. SLAP lesions of the shoulder. A Arthroscopy 1990;6:274–279.
40. Woertler K, Waldt S. MR imaging in sports-related glenohumeral instability. Eur Radiol 2006;16:2622-2636.
41. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psycholog Bull 1968;70(4):213-220.
42. Altman DG. Statistics in medical journals: developments in the 1980s. Stat Med 1991;10(12):1897-1913.

Chapter | 6

Are radiologists superior to orthopaedic surgeons in diagnosing instability related shoulder lesions on magnetic resonance arthrography? A multicenter reproducibility and accuracy study

Susan van Grinsven, Thijs Nijenhuis, Peer Konings,
Albert van Kampen, Corné van Loon

Journal of Shoulder and Elbow Surgery 2015;24(9):1405-1412

Abstract

Introduction: We compared the diagnostic reproducibility and accuracy of musculoskeletal radiologists with orthopaedic shoulder surgeons in 2 large medical centers in assessing magnetic resonance arthrograms (MRA's) of patients with traumatic anterior shoulder instability.

Materials and methods: Forty-five surgically confirmed MRA's were assessed by 4 radiologists, 4 orthopaedic surgeons, 2 radiologic teams, and 2 orthopaedic teams. During MRA assessment and surgery, the same 7-lesion scoring form was used. Kappa coefficients, sensitivity, specificity, and differences in percentage of agreement or correct diagnosis ($p < 0.05$, McNemar test) were calculated per lesion and overall per the 7 lesion types.

Results: The overall kappa between the individual radiologists ($k = 0.51$, $k = 0.46$) and orthopaedic surgeons ($k = 0.46$, $k = 0.41$) was moderate. Although the overall percentage of agreement between the radiologists was slightly higher than that between the orthopaedic surgeons in both centers (80.0% versus 77.5% and 75.2% versus 73.7%), there was no significant difference. In each medical center, however, the most experienced orthopaedic surgeon was exceedingly more accurate than both radiologists per the 7 lesion types (81.9% versus 72.4% / 74.6% and 76.5% versus 67.3% / 73.7%). In 3 of 4 cases, this difference was significant. Overall accuracy improvement through consensus assessment was merely established for the weakest member of each team.

Conclusion: Experienced orthopaedic surgeons are more accurate than radiologists in assessing traumatic anterior shoulder instability related lesions on MRA. In case of diagnosis disagreement, these orthopaedic surgeons should base their treatment decision on their own MRA interpretation.

Introduction

To restore traumatic anterior shoulder instability (TASI), surgery is often needed because the recurrence rate in young patients is 80% to 90%. [1,21,30,31,38-41] Although orthopaedic surgeons base their diagnosis on patient history, physical examination findings, and plain radiographs [5,23,24,31,40-42], additional magnetic resonance arthrography (MRA) imaging is used to confirm the clinical diagnosis. [1,3,5,6,9-12,15-17,19-21,23,25, 28-31,33-37,39,41]

The impact of MRA assessment by radiologists is high because imaging results (lesion type, location, and severity) may directly influence the treatment decision of the orthopaedic surgeon from a nonsurgical to arthroscopic or open surgical approach, and treatment success depends on the accuracy of the preoperative MRA diagnosis and agreement between the two disciplines. [1,3,6,9,12,16,21,30,31,36,39,41] The accuracy rates of radiologists for instability related lesions are high, with sensitivities ranging from 79% to 100% and specificities from 85% to 100% [4,9-12,14-16,18,20,21,23,29,31,35-37], although much lower accuracy rates also have been mentioned. [24,25,28,40,41] Agreement between radiologists about the presence of TASI-related lesions on MRA is, however, highly variable, with k values ranging from poor (-0.003) to almost perfect (0.84). [13,15,20,28-30,38,41] The same seems true regarding MRA agreement between radiologists and orthopaedic surgeons (with k values ranging from -0.007 to 0.86), although only a few authors have reported on this subject. [10,13,28,29]

When faced with disagreement about a MRA diagnosis in daily clinical practice, the orthopaedic surgeon has to decide whether he or she should base the treatment decision on his or her own MRA interpretation or that of the radiologist. The literature is not very helpful regarding this because the results have been inconclusive. Both superior [13] and inferior [24] accuracy rates of radiologists and no difference in accuracy [10,28,29] between radiologists and orthopaedic surgeons have been described.

Thus, we prospectively compared the diagnostic reproducibility and accuracy of experienced musculoskeletal radiologists with those of experienced orthopaedic shoulder surgeons assessing high-field MRA's of patients with TASI in 2 large medical centers. Furthermore, we evaluated the influence of consensus assessment compared with individual assessment. We hypothesized that the MRA assessments of the radiologists would be superior and that consensus assessment would outperform individual assessment.

Materials and methods

To perform a prospective diagnostic performance study, we designed the study protocol before data collection was started. No Dutch regional ethics committee was required for this study, and the need for informed consent was waivered. The study was, however, approved by the scientific committee (LTC Alysis Zorggroep) of the Rijnstate hospital

(LTC479/270807) and the radiologic and orthopaedic departments were instructed accordingly. All consecutive patients with suspected TASI who were referred for an 1.5 Tesla MRA scan between 2007 and 2011, after undergoing TASI reduction and stabilizing surgery, were considered for enrolment. Patients who underwent previous shoulder surgery, who underwent a shoulder procedure between magnetic resonance imaging (MRI) and stabilizing surgery, who had epilepsy, or who were skeletally immature (aged < 16 years) were excluded. Ultimately, 45 MRA's were included.

MRA (index test)

Shoulder MRA was performed per protocol at the Rijnstate Hospital (Arnhem, the Netherlands) with a Siemens Magnetom Avanto Tim system [32 x 8], with a small-extremity coil (Siemens AG, Munich, Germany) or an ACS Intera Gyroscan system (replaced in 2008 by an Achieva system) with a Synergy Flex-M surface shoulder coil (Philips, Best, the Netherlands). The sequences used have been fully described in an article we previously published.[41] The patient was positioned supine with the arm slightly abducted and externally rotated. During the abduction external rotation (ABER) view, the patient placed the hand of the affected extremity behind the head or neck. Arthrography was performed within 30 minutes of MRI to obtain optimum imaging. With fluoroscopic guidance, through an anterior approach, a 21-gauge needle was inserted into the inferior or superior-medial quadrant of the humeral head. Correct intra-articular needle position confirmation was obtained by injecting 2 to 3 ml of iodinated contrast agent and 14 to 16 ml of diluted gadolinium complex (Xenetix 300 and Artirem; Guerbet Nederland, Gorinchem, the Netherlands) into the glenohumeral joint. The radiologic reports for MRA's of sufficient quality were submitted to the orthopaedic department to confirm diagnosis and to plan treatment.

Reference standard

All 45 shoulder stabilizations were performed at the Rijnstate Hospital by experienced shoulder surgeons. MRA's were accessible during surgery. Patients received regional anaesthesia with an interscalene brachial plexus block and were placed in the lateral decubitus or beach-chair position. After instability severity and direction testing, the affected arm was slightly abducted and externally rotated under gentle longitudinal traction. A 30º, 4-mm arthroscope was inserted through the posterior portal, and the other instruments were inserted through 2 anterior portals. The shoulder was systematically inspected according to a standardized scoring list that served as the reference standard: greater humeral tuberosity fracture (none, present), cuff lesion (none, partial tear, full-thickness tear), humeral head lesion (none, degeneration, Hill-Sachs lesion), anterior-inferior glenoid lesion (none, degeneration, defect/bony Bankart lesion), anterior-inferior

Figure 1: Example of disagreement about a cuff lesion (oblique coronal orientation, T1 weighted). The surgical outcome showed no cuff tear, whereas the outcome of magnetic resonance arthrography assessment ranged from normal findings to a partial cuff tear. The arrow on the magnetic resonance arthrogram indicates some irregularity at the footprint of the articular side of the supraspinatus muscle

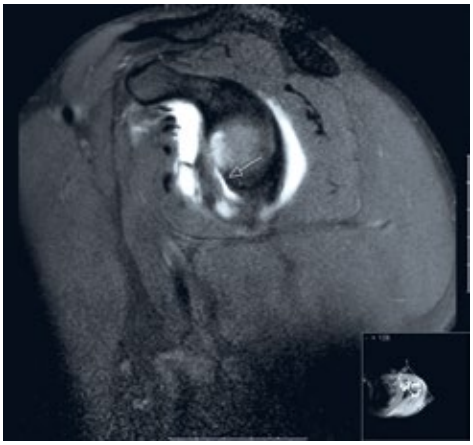

Figure 2: Example of disagreement about an anterior-inferior glenoid lesion (sagittal orientation, T1 weighted). The surgical outcome showed no bony Bankart lesion, whereas the outcome of magnetic resonance arthrography assessment ranged from normal findings to a bony Bankart lesion. The arrow on the magnetic resonance arthrogram indicates a slight suggestion of a bony component situated at the anterior-inferior rim of the glenoid

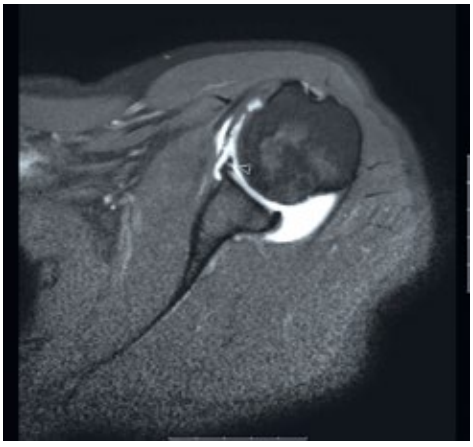

Figure 3: Example of disagreement about an anterior-inferior labrum lesion (axial orientation, T1 weighted). The surgical outcome showed no Bankart lesion, whereas the outcome of magnetic resonance arthrography assessment ranged from normal findings to a Bankart lesion. The arrowhead on the magnetic resonance arthrogram indicates irregularity of the anterior-inferior labrum without separation from the anterior-inferior rim of the glenoid

Figure 4: Example of disagreement about a superior labrum anterior-to-posterior (SLAP) lesion (axial orientation, T1 weighted). The surgical outcome showed a SLAP lesion, whereas the outcome of magnetic resonance arthrography assessment ranged from normal findings to a SLAP lesion. The arrowhead on the magnetic resonance arthrogram indicates similarity with a small sublabral recess (smooth margins)

labrum lesion (none, degeneration/fraying, Bankart lesion), superior labrum anterior-to-posterior (SLAP) lesion (none, SLAP type II-IV), and joint capsule lesion (none, capsule redundancy, anterior stripping/tear) (Figs. 1-4). [7,22,26,27,32,39,43]

Image evaluation

Four radiologists (R1, R2, r1, and r2) and four orthopaedic surgeons (O1, O2, o1, and o2), from 2 different medical centers, evaluated 45 MRA's twice. Medical center X (the Rijnstate Hospital) is a large teaching hospital, offering medical care to 425.000 inhabitants of the Netherlands, with 750 beds. Its radiologists and orthopaedic surgeons had 16 years, 5 years, 8 years, and 6 years of experience. Medical center Y (Sint Maartenskliniek, Nijmegen, the Netherlands) is a large teaching hospital, specializing in gait- and movement-related disorder treatment, with a nationwide adherence and 233 beds. Its radiologists and orthopaedic surgeons had 8 years, 6 years, 6 years, and 3 years of experience.

The first individual MRA assessment series was used to evaluate the reproducibility between observers and the accuracy of each observer per medical center. During the second MRA assessment series, the observers from each medical center were paired per discipline (Team R, Team O, Team r, and Team o) to evaluate the influence of consensus assessment on diagnostic performance. The 7 lesions were assessed with the same scoring list as was used during surgery. The 2 series took place under comparable circumstances at medical center X (e.g. room, time of day, monitor hardware and software) and were separated by a 2-month interval (MRA recognition avoidance). Each series was divided into 2 meetings (fatigue avoidance) with MRA's presented in random order. All observers were provided with a list of all lesion types to create definition uniformity but were blinded to the surgical reports and initial MRA reports.

Statistical analysis

MRA reproducibility and MRA accuracy of the individual observers and teams, as well as between the individual observers and teams, were expressed per medical center, per lesion, and overall per the 7 lesion types using kappa coefficients, sensitivity, specificity and percentage of agreement and percentage of correct diagnosis values. Kappa values were categorized as indicating poor (0.00 to 0.20), fair (0.21 to 0.40), moderate (0.41 to 0.60), substantial (0.61 to 0.80), or almost perfect (0.81 to 1.00) agreement. [8] Differences in percentage of agreement and correct diagnosis were tested for significance with the McNemar test for paired proportions ($P \leq 0.05$). [2] The 7 lesions were dichotomized to construct 2 x 2 tables to facilitate calculations. No adjustment for multiple testing was made. Post hoc sample size calculations indicated that 40 MRA's were required to statistically detect, with a power of 80% and type I error of 5%, a 15% difference in correctly diagnosed lesions between the observers when the proportion of discordant pairs was conservatively assumed to be 0.65 and the method of analysis was the McNemar test. [41] Statistical evaluations were carried out using SPSS software, version 20.0 (IBM, Armonk, NY, USA).

Results

The final study population consisted of 45 surgery confirmed MRA's (31 from male patients and 14 from female patients). The mean time from MRA to surgery was 4.9 months (range, 1 to 24 months). The mean age at the time of surgery was 29.6 years (range, 17 to 57 years). For 10 MRA's, 1 of 5 sequences was missing (7 ABER, 2 T1 oblique coronal, and 1 sagittal). In addition, 2 MRA's were affected by suboptimal contrast agent administration and 7 ABER sequences by motion artefacts. All arthroscopic results and all corresponding data of the 8 individual observers and 4 teams are complete.

MRA reproducibility between individual radiologists and orthopaedic surgeons

The overall k between the 4 observers from center X was moderate (k = 0.42 to 0.51) and between those from center Y was fair (k = 0.36) to moderate (k = 0.49). Per the 7 lesion types, the percentage of agreement between the 4 observers from center X ranged from 76.8% to 80.0% and between those from center Y was slightly lower (71.4% to 76.2%). For center X, the overall k and percentage of agreement between the 2 radiologists (k = 0.51 and 80.0%) were higher than those between the 2 orthopaedic surgeons (k = 0.46 and 77.5%). The same was true for center Y (k = 0.46 and 75.2% versus k = 0.41 and 73.7%). The differences in diagnostic reproducibility between the 2 disciplines were, however, not statistically significant in both centers (Tables 1 and 2).

Table 1: Reproducibility of MRA diagnosis between individual observers or between teams per lesion type and overall per 7 lesion types (N=45: 315 pair-wise ratings)

	Medical Center X			
	R1 / O1	***R1 / O2***	***R2 / O1***	***R2 / O2***
Cuff lesion‡	0.44 (84.4)	0.39 (82.2)	0.41 (80.0)	0.36 (77.8)
Humeral head lesion‡	0.46 (73.3)	0.37 (68.9)	0.59 (82.2)	0.59 (82.2)
Anterior inferior glenoid lesion‡	0.48 (95.6)	0.66 (97.8)	0.79 (97.8)	0.48 (95.6)
Greater humeral tuberosity fractures‡	NA (97.8)	NA (97.8)	NA (97.8)	NA (97.8)
Anterior inferior labrum lesion‡	0.21 (53.3)	0.15 (55.6)	0.14 (71.1)	0.03 (55.6)
SLAP lesion‡	0.22 (86.7)	0.28 (84.4)	-0.04 (62.2)	0.30 (73.3)
Joint capsule lesion‡	0.26 (57.8)	-0.01 (68.9)	0.09 (46.7)	0.04 (75.6)
Overall absolute % of agreement∞	**78.4**	**79.4**	**76.8**	**79.7**
Overall kappa (95% CI)∞	**0.474 (0.372 - 0.576)**	**0.421 (0.294 - 0.530)**	**0.482 (0.380 - 0.584)**	**0.508 (0.406 - 0.610)**

	Medical Center X		
	R1 / R2	***O1 / O2***	***TR / TO***
Cuff lesion‡	0.75 (91.1)	0.44 (84.4)	0.34 (86.7)
Humeral head lesion‡	0.55 (77.8)	0.48 (77.8)	0.65 (84.4)
Anterior inferior glenoid lesion‡	0.37 (93.3)	0.66 (97.8)	NA (95.6)
Greater humeral tuberosity fractures‡	NA (100.0)	-0.02 (95.6)	NA (100.0)
Anterior inferior labrum lesion‡	0.23 (55.6)	0.16 (62.2)	0.30 (73.3)
SLAP lesion‡	0.17 (66.7)	-0.04 (88.9)	-0.04 (73.3)
Joint capsule lesion‡	0.32 (75.6)	-0.04 (35.6)	0.30 (84.4)
Overall absolute % of agreement∞	**80.0**	**77.5**	**85.4**
Overall kappa (95% CI)∞	**0.507 (0.405 - 0.609**	**0.460 (0.356 - 0.564)**	**0.622 (0.524 - 0.720)**

	Medical Center Y			
	r1 / o1	***r1 / o2***	***r2 / o1***	***r2 / o2***
Cuff lesion‡	0.25 (71.1)	0.17 (64.4)	0.32 (75.6)	0.35 (73.3)
Humeral head lesion‡	0.72 (86.7)	0.54 (75.6)	0.73 (86.7)	0.35 (66.7)
Anterior inferior glenoid lesion‡	0.24 (82.2)	0.29 (91.1)	0.09 (80.0)	-0.06 (88.9)
Greater humeral tuberosity fractures‡	NA (91.1)	0.54 (93.3)	NA (97.8)	0.48 (95.6)
Anterior inferior labrum lesion‡	0.29 (62.2)	0.03 (53.3)	0.24 (62.2)	0.15 (57.8)
SLAP lesion‡	0.47 (75.6)	0.36 (68.9)	0.39 (71.1)	0.27 (64.4)
Joint capsule lesion‡	-0.08 (46.7)	0.10 (53.3)	0.18 (60.0)	0.24 (57.8)
Overall absolute % of agreement∞	**73.7**	**71.4**	**76.2**	**72.1**
Overall kappa (95% CI)∞	**0.433 (0.331 - 0.535)**	**0.388 (0.278 - 0.498)**	**0.490 (0.390 - 0.590)**	**0.357 (0.249 - 0.465)**

Table 1: Continued

	Medical Center Y		
	r1 / r2	***o1 / o2***	***tr / to***
Cuff lesion‡	0.25 (68.9)	0.37 (75.6)	0.13 (71.1)
Humeral head lesion‡	0.64 (82.2)	0.38 (66.7)	0.34 (66.7)
Anterior inferior glenoid lesion‡	-0.08 (84.4)	0.35 (86.7)	0.66 (97.8)
Greater humeral tuberosity fractures‡	0.38 (93.3)	NA (93.3)	0.66 (97.8)
Anterior inferior labrum lesion‡	0.29 (64.4)	0.10 (46.7)	0.05 (48.9)
SLAP lesion‡	0.53 (77.8)	0.68 (84.4)	0.48 (75.6)
Joint capsule lesion‡	0.05 (55.6)	0.26 (62.2)	0.00 (48.9)
Overall absolute % of agreement∞	**75.2**	**73.7**	**72.4**
Overall kappa (95% CI)∞	**0.457**	**0.411**	**0.359**
	(0.355 - 0.559)	**(0.307 - 0.515)**	**(0.249 - 0.469)**

N: number of included patients, %: percentage, CI: confidence interval, NA: not applicable (calculation of kappa value not possible due to empty cells in 2x2 table), MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, R1 / r1: radiologist 1, R2/ r2: radiologist 2, O1/ o1: orthopaedic surgeon 1, O2 / o2: orthopaedic surgeon 2, TO / To: team orthopaedic surgeons, TR / Tr: team radiologists ∞: overall value per 7 lesion types, ‡: κ-coefficients (absolute percentage of agreement)

Table 2: Differences in percentages absolute MRA agreement, per lesion type and overall per 7 lesion types between paired individual radiologists and orthopaedic surgeons or between paired teams and their consisting paired individual radiologists or orthopaedic surgeons, tested using the McNemar test for paired proportions (p-values, N=45: 315 pair-wise ratings)

	Medical Center X		
	R1R2 / O1O2	***TRTO / R1R2***	***TRTO / O1O2***
Cuff lesion	0.453	0.754	1.000
Humeral head lesion	1.000	0.581	0.549
Anterior inferior glenoid lesion	0.500	1.000	1.000
Greater humeral tuberosity fractures	NA	NA	NA
Anterior inferior labrum lesion	0.626	0.096	0.359
SLAP lesion	0.013*	0.581	0.065
Joint capsule lesion	0.000*	0.424	0.000*
Overall absolute % of agreement∞	**0.434**	**0.060**	**0.009***
Which observer scores best?	**R1R2**	**TRTO**	**TRTO**
	Medical Center Y		
	r1r2 / o1o2	***trto / r1r2***	***trto / o1o2***
Cuff lesion	0.648	1.000	0.774
Humeral head lesion	0.118	0.143	1.000
Anterior inferior glenoid lesion	1.000	0.070	0.125
Greater humeral tuberosity fractures	1.000	0.625	0.500
Anterior inferior labrum lesion	0.134	0.118	1.000
SLAP lesion	0.581	1.000	0.388
Joint capsule lesion	0.690	0.607	0.238
Overall absolute % of agreement∞	**0.699**	**0.407**	**0.752**
Which observer scores best?	**r1r2**	**r1r2**	**o1o2**

*N: number of included patients, %: percentage, NA: not applicable (calculation of kappa value not possible due to empty cells in 2x2 table), MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, R1 / r1: radiologist 1, R2/ r2: radiologist 2, O1/ o1: orthopaedic surgeon 1, O2 / o2: orthopaedic surgeon 2, TO / to: team orthopaedic surgeons, TR / tr: team radiologists, ∞: overall value per 7 lesion types, * = Significant difference (p < 0.05)*

Table 3: Accuracy (% Sens, Spec and Acc) of MRA diagnoses of individual radiologists and orthopaedic surgeons per lesion type and overall per 7 lesions types (N=45, 315 pair-wise rating)

				Medical Center X		
	R1			***R2***		
	Sens	Spec	Acc	Sens	Spec	Acc
Cuff lesion	80.0	90.0	88.9	80.0	80.0	80.0
Humeral head lesion	57.1	58.8	57.8	78.6	47.1	66.7
Anterior inferior glenoid lesion	0	95.2	88.9	66.7	97.6	95.6
Greater humeral tuberosity fractures	NA	100	100	NA	100	100
Anterior inferior labrum lesion	39.5	85.7	46.7	78.9	42.9	73.3
SLAP lesion	50.0	86.0	84.4	50.0	65.1	64.4
Joint capsule lesion	25.8	71.4	40.0	22.6	85.7	42.2
Overall Sens, Spec, and Acc	**41.1**	**88.5**	**72.4**	**61.7**	**81.3**	**74.6**
				Medical Center X		
	O1			***O2***		
	Sens	Spec	Acc	Sens	Spec	Acc
Cuff lesion	60.0	90.0	86.7	60.0	87.5	84.4
Humeral head lesion	71.4	35.3	57.8	71.4	35.3	57.8
Anterior inferior glenoid lesion	33.3	97.6	93.3	0	97.6	91.1
Greater humeral tuberosity fractures	NA	97.8	97.8	NA	97.8	97.8
Anterior inferior labrum lesion	89.5	57.1	84.4	65.8	85.7	68.9
SLAP lesion	0	97.7	93.3	0	90.7	86.7
Joint capsule lesion	67.7	42.9	60.0	9.7	92.6	35.6
Overall Sens, Spec, and Acc	**73.8**	**86.1**	**81.9**	**47.7**	**88.5**	**74.6**
				Medical Center Y		
	r1			***r2***		
	Sens	Spec	Acc	Sens	Spec	Acc
Cuff lesion	80.0	75.0	75.6	80.0	80.0	80.0
Humeral head lesion	60.7	47.1	55.6	57.1	52.9	55.6
Anterior inferior glenoid lesion	0	90.5	84.4	33.3	95.2	91.1
Greater humeral tuberosity fractures	NA	91.1	91.1	NA	97.8	97.8
Anterior inferior labrum lesion	47.4	71.4	51.1	55.3	85.7	60.0
SLAP lesion	50.0	65.1	64.4	100	62.8	64.4
Joint capsule lesion	54.8	35.7	48.9	77.4	42.9	66.7
Overall Sens, Spec, and Acc	**53.3**	**74.5**	**67.3**	**63.6**	**78.8**	**73.7**

Table 3: Continued

	Medical Center Y					
	o1			**o2**		
	Sens	Spec	Acc	Sens	Spec	Acc
Cuff lesion	60.0	85.0	82.2	60.0	72.5	71.1
Humeral head lesion	71.4	52.9	64.4	42.9	82.4	57.8
Anterior inferior glenoid lesion	66.7	80.0	84.4	33.3	97.6	93.3
Greater humeral tuberosity fractures	NA	100	100	NA	93.3	93.3
Anterior inferior labrum lesion	84.2	57.1	80.0	34.2	71.4	40.0
SLAP lesion	0	60.5	57.8	0	51.1	51.1
Joint capsule lesion	64.5	71.4	66.7	41.9	71.4	51.1
Overall Sens, Spec, and Acc	**72.0**	**78.8**	**76.5**	**39.3**	**78.8**	**65.4**

N: number of included patients, %: percentage, NA: not applicable since fractures of the greater humeral tubercle were not observed, SLAP: superior labral anterior-to-posterior lesion, Sens: sensitivity, specificity, Acc: percentage of correct diagnosis values calculated as: (true-positive + true-negative) / total patients in 2x2 table, R1 / r1: radiologist 1, R2 / r2: radiologist 2, O1 / o1: orthopaedic surgeon 1, O2 / o2: orthopaedic surgeon 2

MRA accuracy between individual radiologists and orthopaedic surgeons

The overall sensitivity and specificity of the 4 observers from center X (41.1% to 73.8% and 81.3% to 88.5%, respectively) were slightly higher than those of the 4 observers from center Y (39.3% to 72.0% and 74.5% to 78.8%, respectively). Overall, the most experienced orthopaedic surgeon from center Y was more accurate (76.5%) than both radiologists (73.7% [$p = 0.356$] and 67.3% [$p = 0.002$]). An even stronger pattern of superior diagnostic accuracy was established by the orthopaedic surgeons from center X: The most experienced orthopaedic surgeon (81.9%) was significantly more accurate, per the 7 lesion types, than both his radiologic colleagues (72.4% [$p < 0.001$] and 74.6% [$p = 0.010$]), and the lesser experienced orthopaedic surgeon (74.6%) was at least equally accurate (72.4% and 74.6%). Per lesion type, the orthopaedic surgeons were especially more accurate in the case of anterior-inferior labrum and SLAP lesions (Tables 3 and 4).

Influence of consensus assessment (teams) on MRA reproducibility

In center Y, the overall k and percentage of agreement between the 2 teams were not statistically different than those between the 2 individual radiologists or orthopaedic surgeons. The overall k between the radiologic team and orthopaedic team from center X, however, showed substantial agreement ($k = 0.62$) and was exceedingly higher than the moderate agreement between the 2 individual radiologists ($k = 0.51$) or orthopaedic

Table 4: Differences in percentages correctly diagnosed per lesion type and overall per 7 lesion types between individual observers, tested using McNemar test for paired proportions (p-values, N=45: 315 pair-wise ratings)

			Medical Center X			
	R1 / R2	***R1 / O1***	***R1 / O2***	***R2 / O1***	***R2 / O2***	***O1 / O2***
Cuff lesion	0.125	1.000	0.727	0.508	0.754	1.000
Humeral head lesion	0.344	1.000	1.000	0.289	0.289	1.000
Anterior inferior glenoid lesion	0.250	0.500	1.000	1.000	0.500	1.000
Greater humeral tuberosity fractures	NA	NA	NA	NA	NA	1.000
Anterior inferior labrum lesion	0.012*	0.000*	0.041*	0.267	0.824	0.143
SLAP lesion	0.035*	0.219	1.000	0.002*	0.006*	0.375
Joint capsule lesion	1.000	0.064	0.791	0.152	0.549	0.061
Overall, per 7 lesion types	**0.450**	**0.000***	**0.457**	**0.010***	**1.000**	**0.009***
Which observer scores best?	**R2**	**O1***	**O2**	**O1***	**O2=R2**	**O1***
			Medical Center Y			
	r1 / r2	***r1 / o1***	***r1 / o2***	***r2 / o1***	***r2 / o2***	***o1 / o2***
Cuff lesion	0.791	0.581	0.804	1.000	0.388	0.227
Humeral head lesion	1.000	0.219	1.000	0.219	1.000	0.607
Anterior inferior glenoid lesion	0.453	1.000	0.125	0.508	1.000	0.219
Greater humeral tuberosity fractures	0.250	NA	1.000	NA	0.500	NA
Anterior inferior labrum lesion	0.454	0.002*	0.383	0.049*	0.064	0.000*
SLAP lesion	1.000	0.549	0.180	0.581	0.210	0.453
Joint capsule lesion	0.115	0.152	1.000	1.000	0.167	0.143
Overall, per 7 lesion types	**0.031***	**0.002***	**0.598**	**0.356**	**0.007***	**0.000***
Which observer scores best?	**r2***	**o1***	**r1**	**o1**	**r2***	**o1***

*N: number of included patients, NA: not applicable since fractures of the greater humeral tubercle were not observed, SLAP: superior labral anterior-to-posterior lesion, R1/ r1: radiologist 1, R2 / r2: radiologist 2, O1 / o1: orthopaedic surgeon 1, O2 / o2: orthopaedic surgeon 2, * = Significant difference (p < 0.05)*

surgeons (k = 0.46). Per the 7 lesion types, the percentage of agreement between the 2 teams from center X was superior to that of the 2 individual radiologists (85.4% versus 80.0%, p = 0.060) and significantly higher than the reproducibility between the 2 individual orthopaedic surgeons (85.4% versus 77.5%, p = 0.009) (Tables 1 and 2).

Table 5: Accuracy (% Sens, Spec and Acc) of MRA diagnosis of Teams radiologists and Teams orthopaedic surgeons, per lesion type and overall per 7 lesions types (N=45: 315 pair-wise rating)

			Medical Center X			
	Team R			***Team O***		
	Sens	Spec	Acc	Sens	Spec	Acc
Cuff lesion	80.0	92.5	91.1	40.0	97.5	91.1
Humeral head lesion	78.6	41.2	64.4	71.4	47.1	62.2
Anterior inferior glenoid lesion	0	95.2	88.9	0	100	93.3
Greater humeral tuberosity fractures	NA	100	100	NA	100	100
Anterior inferior labrum lesion	73.7	57.1	71.1	89.5	57.1	84.4
SLAP lesion	100	79.1	80.0	0	97.7	93.3
Joint capsule lesion	16.1	78.6	35.6	6.5	92.9	33.3
Overall Sens, Spec, and Acc	**57.0**	**85.6**	**75.9**	**54.2**	**92.8**	**79.7**
			Medical Center Y			
	Team r			***Team o***		
	Sens	Spec	Acc	Sens	Spec	Acc
Cuff lesion	60.0	85.0	82.2	40.0	80.0	75.6
Humeral head lesion	85.7	23.5	62.2	60.7	70.6	64.4
Anterior inferior glenoid lesion	0	97.6	91.1	0	95.2	88.9
Greater humeral tuberosity fractures	NA	95.6	95.6	NA	97.8	97.8
Anterior inferior labrum lesion	34.2	85.7	42.2	68.4	85.7	71.1
SLAP lesion	0	55.8	53.3	0	67.4	64.4
Joint capsule lesion	61.3	64.3	62.2	35.3	71.4	46.7
Overall Sens, Spec, and Acc	**55.1**	**77.4**	**69.8**	**52.3**	**83.2**	**72.7**

N: number of included patients, %: percentage, NA: not applicable (accuracy rates could not be determined due to empty cells in 2x2 table), SLAP: superior labral anterior-to-posterior lesion, Sens: sensitivity, Spec: specificity, Acc: percentage of correct diagnosis values calculated as: (true-positive + true-negative) / total patients in 2x2 table, Team O / Team o: team orthopaedic surgeons, Team R / Team r: team radiologists

Influence of consensus assessment (teams) on MRA accuracy

In both centers, the overall accuracy of the orthopaedic team was higher than that of the radiologic team (79.7% versus 75.9% and 72.7% versus 69.8%). Both orthopaedic teams significantly improved the overall diagnostic accuracy of their weakest orthopaedic surgeon (79.7% versus 74.6% [p = 0.023] and 72.7% versus 65.4% [p = 0.009]). The same pattern was established by the radiologic teams, though not significantly. The radiologic team from center Y enhanced the overall accuracy of its weakest radiologist (69.8% versus

Table 6: Differences in percentages correctly diagnosed per lesion type and overall per 7 lesion types between the teams and between the teams and their individual observers, tested using McNemar test for paired proportions (p-values, N=45: 315 pair-wise ratings)

	Medical Center X				
	TR / TO	***TR / R1***	***TR / R2***	***TO / O1***	***TO / O1***
Cuff lesion	1.000	1.000	0.180	0.687	0.453
Humeral head lesion	1.000	0.549	1.000	0.727	0.754
Anterior inferior glenoid lesion	0.500	1.000	0.250	1.000	1.000
Greater humeral tuberosity fractures	NA	NA	NA	NA	NA
Anterior inferior labrum lesion	0.146	0.013*	1.000	1.000	0.118
SLAP lesion	0.146	0.727	0.039*	1.000	0.375
Joint capsule lesion	1.000	0.774	0.607	0.029*	1.000
Overall, per 7 lesion types	**0.104**	**0.177**	**0.672**	**0.410**	**0.023***
Which observer(s) scores best?	**TO**	**TR**	**TR**	**O1**	**TO**
	Medical Center Y				
	Tr / To	***Tr / r1***	***Tr / r2***	***To / o1***	***To / o2***
Cuff lesion	0.581	0.453	1.000	0.581	0.774
Humeral head lesion	1.000	0.581	0.581	1.000	0.549
Anterior inferior glenoid lesion	1.000	0.250	1.000	0.727	0.500
Greater humeral tuberosity fractures	1.000	0.500	1.000	NA	0.500
Anterior inferior labrum lesion	0.011*	0.481	0.057	0.344	0.001*
SLAP lesion	0.227	0.267	0.302	0.375	0.109
Joint capsule lesion	0.210	0.238	0.815	0.064	0.804
Overall, per 7 lesion types	**0.391**	**0.416**	**0.201**	**0.169**	**0.009***
Which observer(s) scores best?	**To**	**Tr**	**r2**	**o1**	**To***

*N: number of included patients, NA: not applicable since fractures of the greater humeral tubercle were not observed, SLAP: superior labral anterior-to-posterior lesion, R1 / r1: radiologist 1, R2 / r2: radiologist 2, O1 / o1: orthopaedic surgeon 1, O2 / o2: orthopaedic surgeon 2, TO / To: team orthopaedic surgeons, TR / Tr: team radiologists, * = Significant difference ($p < 0.05$)*

67.3%), and the radiologic team from center X improved the overall accuracy of both of its radiologists (79.7% versus 72.4% and 79.7% versus 74.6%). Per lesion type, significantly superior accuracy through consensus assessment was especially found in the case of anterior-inferior labrum and SLAP lesions (Tables 5 and 6).

Discussion

In this prospective diagnostic study, we were unable to support our hypothesis that the diagnostic performance of experienced musculoskeletal radiologists, assessing TASI-related MRA's, is superior to that of experienced orthopaedic shoulder surgeons. Per medical center, the overall reproducibility between a radiologist and an orthopaedic surgeon ranged from k = 0.42 to k = 0.51 and from k = 0.36 to k = 0.49. Radiologists agreed just slightly more (80.0% and 75.2%) about the presence of 7 instability related lesions compared with orthopaedic surgeons (77.5% and 73.7%). These differences were not statistically significant. The moderate overall agreement between the two disciplines is situated in the midsection of the range (k = -0.03 to 0.84) mentioned in earlier literature [13,15,20,28-30,38,41] and concurs with the results of 2 studies by Schreinemachers et al. [28,29] They assessed cuff and anterior-inferior labrum lesions on 92 MRA's and found only moderate agreement between 2 musculoskeletal radiologists and an orthopaedic surgeon (k = 0.48 to 0.56 and k = 0.44 to 0.56). Although we provided all observers with a lesion definition list to create uniformity, it seems that radiologists and orthopaedic surgeons frequently differ in the interpretation of what defines instability related shoulder lesions or have a different use of terminology. [13,42]

The overall sensitivity of the individual observers from medical center X was slightly lower (41.1% to 73.8%) than the results in previous literature (79% to 100%) and the overall specificity was comparable (81.3% to 88.5%) with the results in previous literature (85% to 100%). [4,9-12,14-16,18,20,21,23,29,31,35-37] The overall accuracy rates for center Y were lower in general (sensitivity of 39.3% to 72.0% and specificity of 74.5% to 78.8%), probably because its radiologists and orthopaedic surgeons had to adjust to slight hardware, software, and sequence differences during the MRA assessments in medical center X. The overall percentage of correct diagnosis of both centers (65.4% to 81.9%), however, concurs perfectly with our previous retrospective study in which the same 7 lesions were assessed on 61 MRA's and accuracy of 65.4% to 78.9% was found. [41]

Per the 7 lesion types, 3 of 4 orthopaedic surgeons (O1, 81.9%; O2, 74.6%; and o2, 76.5%) were equally as accurate as or significantly more accurate than their radiologic colleagues (R1, 72.4%; R2, 74.6%; r1, 67.3%; and r2, 73.7%). The orthopaedic surgeons were especially more accurate in the case of anterior-inferior labrum and SLAP lesions. Our results confirm those in a study by Pandya et al. [24], in which an orthopaedic surgeon (sensitivity, 72%) was also significantly more accurate in assessing SLAP lesions on 51 MRA's than a radiologist (sensitivity, 50%). We agree with their explanation that the superior accuracy of orthopaedic surgeons might be caused by suspicion for SLAP or anterior-inferior lesions based on clinical diagnosis (patient history and physical examination findings). Furthermore, in contradiction to radiologists, orthopaedic surgeons are provided with instant feedback during stabilizing surgery. This discrepancy in personal feedback and focus between orthopaedic surgeons (focusing specifically on instability

related lesions) and radiologists (focusing on every abnormality or defect) could have diverged the professional development of the 2 disciplines, causing the higher accuracy of orthopaedic surgeons.

The only exception was orthopaedic surgeon o2 from medical center Y (65.4%), who was equally as accurate as radiologist r1 (67.3%) but significantly less accurate than radiologist r2 (73.7%, $p = 0.007$). He was, however, the least experienced of all observers (3 years versus $\geq$ 5 years of experience). The negative influence of low experience on diagnostic accuracy is supported by former literature. [10,19,25,31,35,38,41] Dinter et al. [10] compared the diagnostic MRA performance of a radiology fellow with that of an experienced musculoskeletal radiologist and an experienced orthopaedic shoulder surgeon in the assessment of 29 patients with cuff, labral, and SLAP lesions. The experienced assessors were significantly more sensitive (67% to 92.9%) than the inexperienced assessor (11% to 50%). Theodoropoulos et al. [35] compared the diagnostic MRA performance of musculoskeletal radiologists with general radiologists in the assessment of 250 labral, rotator cuff, biceps, and Hill-Sachs lesions. The musculoskeletal radiologists were significantly more sensitive (75% to 87% versus 20% to 60%).

We discovered a firm trend showing that consensus assessment of radiologists (i.e., radiologic team) or orthopaedic surgeons (i.e., orthopaedic team) improves the overall accuracy of the weakest team member. Although future research concerning the added value of consensus assessment is necessary, because there is no literature available to confirm our results, it seems fair to suggest that inexperienced radiologists or orthopaedic surgeons should not hesitate to consult with more experienced peers in case of difficult to diagnose instability related lesions on MRA to improve their diagnostic accuracy.

We included only surgery confirmed MRA's in our study. Surgery decision was based on patient history, physical examination findings, radiographs, and the MRA report. As such, (partial) information bias could have been introduced. Furthermore, 4% of the sequences of the MRI protocol were missed (7 ABER, 1 sagittal, and 2 T1 oblique coronal). This could have lowered the diagnostic performance. The circumstances were, however, equal for all observers, representative of normal clinical practice, and in line with the literature. [28,29,31] Finally, stabilizing shoulder surgery is a surgeon-dependent method, although it is the best reference standard available. A strength of this study is that it is the first study, using a consistent protocol and defined lesions to create uniformity among reviewers, that has prospectively compared the diagnostic performance of experienced musculoskeletal radiologists with that of orthopaedic shoulder surgeons in the assessment of TASI related MRA's in 2 large medical centers.

Conclusion

Diagnosis disagreement between radiologists and orthopaedic surgeons frequently occurs in daily practice. Because this can lead to suboptimal treatment decisions, we compared the diagnostic performance of musculoskeletal radiologists with that of orthopaedic shoulder surgeons in the assessment of instability related shoulder MRA's in 2 large medical centers. Furthermore, we evaluated the influence of consensus assessment compared with individual assessment. We hypothesized that the MRA assessments of the radiologists would be superior and that consensus assessment would outperform individual assessment. We indeed discovered a firm trend showing that consensus assessment improves the overall accuracy of the weakest team member. The diagnostic performance of the radiologists was, surprisingly, not superior. The overall agreement between the 2 disciplines about the presence of instability related lesions was not statistically different. Experienced orthopaedic shoulder surgeons are, however, equally as accurate as or significantly more accurate than radiologists in assessing TASI-related lesions on MRA. Therefore, if the clinical diagnosis is not confirmed by the radiologist's MRA report, the orthopaedic surgeon should base his or her treatment decision on his or her own MRA interpretation because treatment success depends on an accurate preoperative diagnosis. Intensified personal feedback after stabilizing surgery and consultation with an experienced peer, in the case of difficult to diagnose instability related lesions on MRA, should be promoted to fine-tune definition interpretation agreement between the 2 disciplines and enhance accuracy. Diagnostic performance improvement, through intensified collaboration of radiologists and orthopaedic surgeons and exchange of knowledge between the 2 disciplines, will ultimately result in better treatment decisions and better health care.

Disclaimer

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References

1. Acid S, Le Corroller T, Aswad R, Pauly V, Champsaur P. Preoperative imaging of anterior shoulder instability: diagnostic effectiveness of MDCT arthrography and comparison with MR arthrography and arthroscopy. AJR Am J Roentgenol 2012;198:661-667.
2. Altman DG. Statistics in medical journals: developments in the 1980s. Stat Med 1991;10: 1897-1913.
3. Amin MF, Youssef AO. The diagnostic value of magnetic resonance arthrography of the shoulder in detection and grading of SLAP lesions: comparison with arthroscopic findings. Eur J Radiol 2012;81:2343-2347.
4. Applegate GR, Hewitt M, Snyder SJ, Watson E, Kwak S, Resnick D. Chronic labral tears: value of magnetic resonance arthrography in evaluating the glenoid labrum and labral-bicipital complex. Arthroscopy 2004;20: 959-963.
5. Bak K, Wiesler ER, Poehling GG. Consensus statement on shoulder instability. Arthroscopy 2010;26:249-255.
6. Bencardino J, Gyftopoulos S, Palmer W. Imaging in anterior glenohumeral instability. Radiology 2013;269:323-37.
7. Blum A, Coudane H, Molé D. Gleno-humeral instabilities. Eur Radiol 2000;10:63-82.
8. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70:213-220.
9. de Jesus JO, Parker L, Frangos AJ, Nazarian LN. Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. AJR Am J Roentgenol 2009;192:1701-1707.
10. Dinter DJ, Martetschläger F, Büsing KA, Schönberg SO, Scharf HP, Lehmann LJ. Shoulder injuries in overhead athletes: utility of MR arthrography. Sportverletz Sportschaden 2008;22: 146-152.
11. Drescher R, von Rothenburg T, Ludwig J, Köster O, Schmid G. Direct MR-arthrography of the shoulder with maximum capsular distension for surgical planning. Rofo 2004;176:1469-1474.
12. Genovese E, Spano E, Castagna A, Leonardi A, Angeretti MG, Callegari L, et al. MR-arthrography in superior instability of the shoulder: correlation with arthroscopy. Radiol Med 2013;118:1022-1033.
13. Halma JJ, Eshuis R, Krebbers YM, Weits T, de Gast A. Interdisciplinary inter-observer agreement and accuracy of MR imaging of the shoulder with arthroscopic correlation. Arch Orthop Trauma Surg 2012;132:311-320.
14. Hayes ML, Collins MS, Morgan JA, Wenger DE, Dahm DL. Efficacy of diagnostic magnetic resonance imaging for articular cartilage lesions of the glenohumeral joint in patients with instability. Skeletal Radiol 2010;39:1199-1204.
15. Holzapfel K, Waldt S, Bruegel M, Paul J, Heinrich P, Imhoff AB, et al. Inter- and intraobserver variability of MR arthrography in the detection and classification of superior labral anterior posterior (SLAP) lesions: evaluation in 78 cases with arthroscopic correlation. Eur Radiol 2010;20:666-673.
16. Iqbal HJ, Rani S, Mahmood A, Brownson P, Aniq H. Diagnostic value of MR arthrogram in SLAP lesions of the shoulder. Surgeon 2010;8:303-309.

17. Ito Y, Sakai T, Tomo H, Nakao Y, Inui K, Koike T, et al. Computerized assessment of Bankart lesions under tension with magnetic resonance arthrography. J Shoulder Elbow Surg 2005;14: 247-251.
18. Jana M, Srivastava DN, Sharma R, Gamanagatti S, Nag HL, Mittal R, et al. Magnetic resonance arthrography for assessing severity of glenohumeral labroligamentous lesions. J Orthop Surg (Hong Kong) 2012;20:230-235.
19. Jonas SC, Walton MJ, Sarangi PP. Is MRA an unnecessary expense in the management of a clinically unstable shoulder? A comparison of MRA and arthroscopic findings in 90 patients. Acta Orthop 2012;83:267-270.
20. Jung JY, Jee WH, Chun HJ, Ahn MI, Kim YS. Magnetic resonance arthrography including ABER view in diagnosing partial-thickness tears of the rotator cuff: accuracy, and inter- and intra-observer agreements. Acta Radiol 2010;51:194-201.
21. Khedr SA, Kassem HM, Azab MA. Comparative study of direct MR arthrography and CT arthrography with arthroscopic correlation in preoperative evaluation of anterior shoulder instability. Egypt J Radiol Nucl Med 2013;44:817-827.
22. Matsen FA, Rockwood CA. Glenohumeral instability. The Shoulder. Matsen FA, Rockwood CA, editors. WB Saunder Company, Philadelphia 1998; 611-745.
23. Oh JH, Kim JY, Choi JA, Kim WS. Effectiveness of multidetector computed tomography arthrography for the diagnosis of shoulder pathology: comparison with magnetic resonance imaging with arthroscopic correlation. J Shoulder Elbow Surg 2010;19:14-20.
24. Pandya NK, Colton A, Webner D, Sennett B, Huffman GR. Physical examination and magnetic resonance imaging in the diagnosis of superior labrum anterior-posterior lesions of the shoulder: a sensitivity analysis. Arthroscopy 2008;24:311-317.
25. Reuss BL, Schwartzberg R, Zlatkin MB, Cooperman A, Dixon JR. Magnetic resonance imaging accuracy for the diagnosis of superior labrum anterior-posterior lesions in the community setting: eightythree arthroscopically confirmed cases. J Shoulder Elbow Surg 2006;15:580-585.
26. Sailer J, Imhof H. Schultergelenkinstabilität. Radiologe 2004;44:578-590.
27. Sanders TG, Morrison WB, Miller MD. Imaging techniques for the evaluation of glenohumeral instability. Am J Sports Med 2000;28:414-434.
28. Schreinemachers SA, van der Hulst VP, Willems WJ, Bipat S, van der Woude HJ. Detection of partial-thickness supraspinatus tendon tears: is a single direct MR arthrography series in ABER position as accurate as conventional MR arthrography? Skeletal Radiol 2009;38:967-975.
29. Schreinemachers SA, van der Hulst VP, Willems WJ, Bipat S, van der Woude HJ. Is a single direct MR arthrography series in ABER position as accurate in detecting anteroinferior labroligamentous lesions as conventional MR arthography? Skeletal Radiol 2009;38:675-683.
30. Simão MN, Nogueira-Barbosa MH, Muglia VF, Barbieri CH. Anterior shoulder instability: correlation between magnetic resonance arthrography, ultrasound arthrography and intraoperative findings. Ultrasound Med Biol 2012;38:551-560.
31. Smith TO, Drew BT, Toms AP. A meta-analysis of the diagnostic test accuracy of MRA and MRI for the detection of glenoid labral injury. Arch Orthop Trauma Surg 2012;132:905-919.
32. Snyder SJ, Karzel RP, Del Pizzo W, Ferkel RD, Friedman MJ. SLAP lesions of the shoulder. Arthroscopy 1990;6:274-279.

33. Song HT, Huh YM, Kim S, Lee SA, Kim SJ, Shin KH, et al. Anteriorinferior labral lesions of recurrent shoulder dislocation evaluated by MR arthrography in an adduction internal rotation (ADIR) position. J Magn Reson Imaging 2006;23:29-35.
34. Stetson WB, Phillips T, Deutsch A. The use of magnetic resonance arthrography to detect partial-thickness rotator cuff tears. J Bone Joint Surg Am 2005;87(2):81-88.
35. Theodoropoulos JS, Andreisek G, Harvey EJ, Wolin P. Magnetic resonance imaging and magnetic resonance arthrography of the shoulder: dependence on the level of training of the performing radiologist for diagnostic accuracy. Skeletal Radiol 2010;39:661-667.
36. Tian CY, Cui GQ, Zheng ZZ, Ren AH. The added value of ABER position for the detection and classification of anteroinferior labroligamentous lesions in MR arthrography of the shoulder. Eur J Radiol 2013;82:651-657.
37. Toyoda H, Ito Y, Tomo H, Nakao Y, Koike T, Takaoka K. Evaluation of rotator cuff tears with magnetic resonance arthrography. Clin Orthop Relat Res 2005;439:109-115.
38. van der Veen HC, Collins JP, Rijk PC. Value of magnetic resonance arthrography in post-traumatic anterior shoulder instability prior to arthroscopy: a prospective evaluation of MRA versus arthroscopy. Arch Orthop Trauma Surg 2012;132:371-375.
39. van der Woude HJ, Vanhoenacker FM. MR arthrography in glenohumeral instability. JBR-BTR 2007;90:377-383.
40. van Grinsven S, Hagenmaier F, van Loon CJ, van Gorp MJ, van Kints MJ, van Kampen A. Does the experience level of the radiologist, assessment in consensus, or the addition of the abduction and external rotation view improve the diagnostic reproducibility and accuracy of MRA of the shoulder? Clin Radiol 2014;69:1157-1164.
41. van Grinsven S, Kesselring FO, van Wassenaer-van Hall HN, Lindeboom R, Lucas C, van Loon CJ. MR arthrography of traumatic anterior shoulder lesions showed modest reproducibility and accuracy when evaluated under clinical circumstances. Arch Orthop Trauma Surg 2007;127:11-17.
42. van Kampen DA, van den Berg T, van der Woude HJ, Castelein RM, Terwee CB, Willems WJ. Diagnostic value of patient characteristics, history, and six clinical tests for traumatic anterior shoulder instability. J Shoulder Elbow Surg 2013;22:1310-1319.
43. Woertler K, Waldt S. MR imaging in sports-related glenohumeral instability. Eur Radiol 2006;16:2622-2636.

Chapter | 7

A feedback protocol improves the diagnostic performance of MR Arthrography by experienced musculoskeletal radiologists in patients with traumatic anterior shoulder instability

Susan van Grinsven, Corné van Loon, Maarten van Gorp,
Marjolijn van Kints, Peer Konings, Albert van Kampen,
Corné van Loon

European Journal of Radiology 2015;84:2242-2249

Abstract

Purpose: To prospectively evaluate the diagnostic performance of magnetic resonance arthrography (MRA) by experienced musculoskeletal radiologists in patients with traumatic anterior shoulder instability (TASI), after feedback protocol execution.
Materials and methods: Forty-five surgically confirmed MRA's were used to enhance personal feedback, to discuss differences in outcome between MRA assessment and surgical findings and to fine-tune definition interpretation agreement of 7 different TASI-related lesions, between experienced musculoskeletal radiologists and experienced orthopaedic shoulder surgeons. After execution of the feedback protocol 20 new, surgically confirmed, MRA's were assessed by 2 experienced musculoskeletal radiologists using a seven-lesion standardized scoring form. Kappa coefficients, sensitivity, specificity and differences in percentage agreement or correct diagnosis (p-value, McNemar test) were calculated per lesion and overall per 7 lesion types to assess whether diagnostic reproducibility and accuracy was improved.
Results: Per 7 lesion types, the overall kappa and percentage of agreement, between the 2 radiologists, were dramatically increased in comparison with our former study (k=0.81 versus k=0.48 and 90.7% versus 78.2%, respectively). The overall sensitivity of radiologist 1 increased from 45.9% to 87.8%, the overall sensitivity of radiologist 2 increased from 63.5% to 79.6% and the overall specificity of radiologist 2 increased from 80.1% to 85.7%. Furthermore, the overall percentage of correct diagnosis of both radiologists was also exceedingly higher (85.7% and 83.6%) compared to our former study (74.4% and 74.8%).
Conclusion: The implementation of our feedback protocol dramatically improved the reproducibility and accuracy of high-field MRA by experienced musculoskeletal radiologists in patients with traumatic anterior shoulder instability.

Introduction

Off all joints in the human body, the shoulder joint is the most mobile one. This advantage in range of motion is however exchanged for a vulnerability to instability. [1-3] In the Netherlands the yearly incidence of shoulder dislocation is estimated at 38 per 100.000 persons. [4] 95% of these dislocations are in anterior direction. [5,6] As the recurrence rate is approximated at 80-90% in young patients [5], anterior shoulder instability is often surgically treated, to correct instability caused by soft tissue (rotator cuff, labroligamentous-complex and capsule) and bony (greater humeral tuberosity, humeral head and glenoid) lesions. [3,6-11] Diagnosis can be strongly suspected by patient history, physical examination and standard x-rays alone [3,4,7,10,12-14], but additional diagnostic imaging is used to support clinical findings (lesion type, location and severity) and to guide treatment decisions (nonsurgical, arthroscopic or open surgical approach). [1,3,5,8-11,15-19]

Currently, magnetic resonance arthography (MRA) is described as the most accurate pre-arthroscopy diagnostic imaging technique for soft-tissue evaluation of the shoulder, being superior to conventional magnetic resonance imaging for the detection of capsular-labroligamentous-complex and rotator cuff lesions [2,3,5,7-9,11,12,15,16,18-27] as the contrast medium distends the joint capsule, outlines intra-articular structures and leaks into tears. [10,11,13,26-28] To increase the diagnostic performance of high-field MRA even more, the use of experienced musculoskeletal radiologists and addition of the abduction external rotation position view sequence (ABER) in the MRA protocol are highly advocated. Authors state that experienced musculoskeletal radiologist perform better than general radiologists [3,6,7,10,20,22,23,28] and ABER improves the visualisation of partial-thickness rotator cuff undersurface tears and lesions of the capsular-labroligamentous complex, through distraction and increased penetration of contract material into a tear. [1,7,9-11,19,24,29-31]

Reviewing the literature, the reported diagnostic MRA reproducibility results of instability related lesions are highly variable with kappa's ranging from -0.03 to 0.84, but most authors describe "moderate" (0.41-0.60) to "substantial agreement" (0.61-080). [3,6,9,21,24,25,30,32] The majority of the authors systematically report high diagnostic MRA accuracy rates for instability related lesions. [2,8,10,13,16-21,25,27,30,33-35] Sensitivities range from 79%-100% and specificities from 85-100% [28], although a few author have mentioned much lower accuracy rates too. [3,7,14,23,24]

If assessment of high-field MRA's, preferably with the additional ABER, is performed by experienced radiologists specializing in musculoskeletal MRA, it should be possible to equal or to increase the substantial reproducibility (kappa: 0.61-0.80) and high accuracy rates (79-100%), mentioned in earlier literature. [2,8-10,13,16-21,24,25,27,28,30,32-35] We were however not able to approximate these results in two former studies. In our first study 2 radiologists retrospectively reviewed 61 high-field MRA's, without ABER. The agreement was "poor" (kappa: 0.21) and although the specificity of the most experienced musculo-

skeletal radiologist was in line with former literature, the sensitivity was only 50%. [3] In the second study we prospectively reviewed 51 MRA's, with ABER. The agreement between our two most experienced musculoskeletal radiologists was "moderate" (kappa: 0.48). Sensitivity ranged from 46-64% and specificity from 80-88%. [7]

As a possible explanation for these disappointing reproducibility and accuracy rates we reasoned that the lack of standard personal feedback in our hospital, addressing the discrepancy between the MRA assessment of the radiologist (focussing on every abnormality or defect, even when it is of no consequence for the stability of the shoulder) and the conclusion of the orthopaedic surgeon after surgical stabilization (focussing entirely on stability related lesions), flattened the learning curve of our radiologists. A high degree of specialized experience could therefore still lead to a suboptimal diagnostic performance. [7] Unfortunately, there is no literature to confirm this. As earlier suggested by other authors, a further explanation could be that radiologists and orthopaedic surgeons interpretate the given definitions of assessed lesions in a different way, consequently leading to lower diagnostic accuracy of the radiologist. [4,32] There are, however, no studies available about how to improve the agreement of definition interpretation.

That is why we developed a feedback protocol in which radiologists systematically received personal feedback of their MRA assessment after surgery, enabling radiologists and orthopaedic surgeons to discuss discrepancies and fine-tune their agreement about lesion definition interpretation. The aim of the present study is to evaluate the diagnostic performance of high-field MRA by experienced musculoskeletal radiologists in patients with traumatic anterior shoulder instability (TASI), after feedback protocol execution. Our hypothesis is that the diagnostic reproducibility and accuracy will improve.

Materials and methods

In order to perform a prospective diagnostic reproducibility and accuracy study, the study protocol was designed before data collection was started. The Regional Ethics Committee of the Netherlands decided that no approval or informed consent was required. The study protocol was, however, approved by the local scientific committee of the Rijnstate hospital and the radiologic / orthopaedic department were instructed accordingly. The Rijnstate hospital is a large teaching hospital with 750 beds and an adherence of 425.000 inhabitants. All patients, visiting our secondary care setting with shoulder instability after reduced TASI, which were referred for an 1.5 Tesla MRA between November 2012 and March 2014 and underwent stabilizing shoulder surgery by orthopaedic shoulder surgeon PK before April 2014, were considered for enrolment. Patients were excluded in case of previous shoulder surgery, record of shoulder procedure between MRI and stabilizing shoulder surgery, skeletal immaturity, general contra-indications for contrast agents and / or MRI or missing ABER-view sequence. Ultimately 20 high-field surgery confirmed MRA's were included.

MRA (index test)

To obtain optimum imaging, the timeframe between arthrography and MRI was less than 30 minutes. According to the radiologists preference, an anterior approach was used to insert a, fluoroscopically guided, 21 gauge needle into the inferior or superior-medial quadrant of the humeral head. Confirmation of correct intra-articular needle position was obtained by injecting 2-3 cc iodinated contrast agent (Xenitix 300, Guerbet Nederland B.V., Gorinchem, the Netherlands) and 14-16 cc diluted gadolinium complex (Artirem, Guerbet Nederland B.V., Gorinchem, the Netherlands) into the glenohumeral joint. Shoulder MRI was performed according to a standardized study protocol (Table 1) with an 1.5 Tesla Achieva system (Philips, Best, the Netherlands) and a Synergy flex-M surface shoulder coil or an 1.5 Tesla Siemens Magnetom Avanto Tim 32x8 (Siemens AG, Munich, Germany) with a Small Extremity coil. Patient placement was supine with the arm slightly in abduction / exorotation. The patient's hand of the affected extremity was placed posterior to the contralateral aspect of the head or neck with a flexed elbow, in case of ABER.

Table 1: MR imaging (1.5 Tesla) protocol

Sequence	***Gradient echo / T1-weighted (T1W/FFE/3D)***	***Turbo spin echo / T2-weighted (T2W/TSE)***	***Proton density weighted / Spin echo (PDW/SE)***	***Turbo spin echo / T1-weighted (T1W/TSE)***	***Turbo spin echo / T1-weighted (T1W/TSE)***
Orientation	Oblique coronal	Oblique coronal	Sagittal	Axial	ABER position[a]
Fat suppressed	Yes	Yes	No	Yes	Yes
Time to repeat (ms)	23	3269	1800	475	475
Time to echo (ms)	9.5	70	25	18	18
Flip angle (°)	20	90	90	90	90
Slice thickness (mm)	4	3.5	3.5	3.5	3.5
Slice gap (mm)	--	0.35	0.35	0.35	0.35
NEX[b]	3	4	2	3	3
Duration (min)	4.10	3.25	4.51	5.28	5.28
FOV[c] (mm)	170	180	130	180	180
Matrix size	304 x 228	256 x 205	288 x 202	304 x 212	256 x 205
Reconstructionmatrix	512	512	512	512	512

[a]: *the optional abduction external rotation view improves visualization of the anterior labroligamentous complex*

[b]: *number of excitations*

[c]: *field of view*

Reference standard

After MRA imaging, the MRA report was send to the orthopaedic department to confirm clinical diagnosis and plan arthroscopic stabilisation. All arthroscopies were performed by the same orthopaedic surgeon (PK), with regional anaesthesia and an additional interscalene brachial plexus block. Before incision, the severity and directions of instability were tested. The patient was placed in either the lateral decubitus position or the beach chair position, with the arm slightly abducted and exorotated. According to protocol, a standard 4-mm 30-degree arthroscope was inserted using the classic posterior approach. Other instruments were inserted via two anterior portals. During arthroscopy, the shoulder was systematically inspected according to a seven-item scoring list rating the presence and severity of instability related lesions. The assessed pathologies were greater humeral

Figure 1 and 2: Example of disagreement about a cuff lesion (oblique coronal orientation, T1 and T2 weighted). Surgical outcome: partial cuff tear. Outcome MRA assessment: ranging from normal to partial cuff tear. Arrows on MRA: difference in outcome probably due to the magic angle phenomenon at the articular side of the m. supraspinatus (T1) and the presence of a tendinosis of the m. Supraspinatus with high signal, apart from the partial cuff tear at the articular side of the m. Supraspinatus (T2)

tuberosity fracture (not present / present), cuff lesion (not present / partial tear / full thickness tear), humeral head lesion (not present / degeneration / Hill-Sachs), anterior-inferior glenoid lesion (not present / degeneration / bony Bankart), anterior-inferior labrum lesion (not present / degeneration / classic Bankart), superior labrum anterior-to-posterior lesion (not present / SLAP II-IV) and joint capsule lesion (not present / capsule redundancy / anterior stripping or tear). [11,36-41] (Figs. 1-6) The surgical scoring list served as the reference standard.

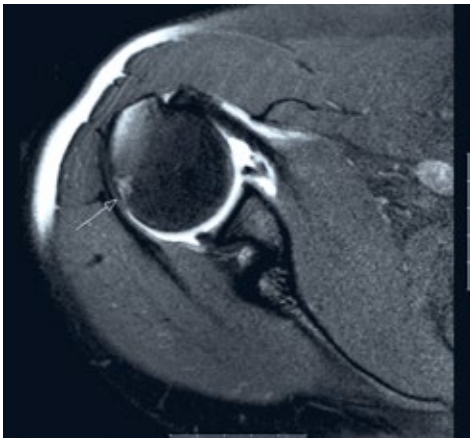

Figure 3: Example of disagreement about a humeral head lesion (axial orientation, T1 weighted): Surgical outcome: Hill-Sachs lesion. Outcome MRA assessment: degeneration. Arrow on MRA: relatively smooth contour of the posterior-lateral side of the humeral head, with a few mini cysts and edema formation

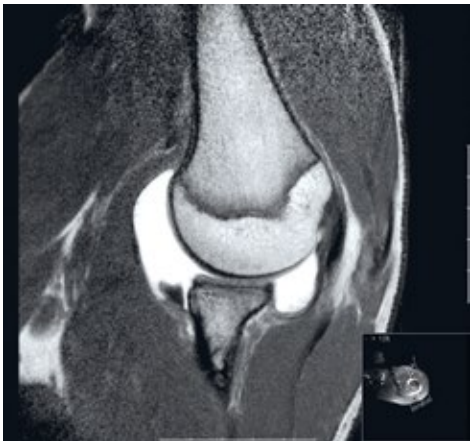

Figure 4: Example of disagreement about an anterior-inferior glenoid lesion (ABER position orientation, T1 weighted). Surgical outcome: no bony Bankart lesion. Outcome MRA assessment: ranging from normal to bony Bankart lesion. Arrow on MRA: flattened aspect of the anterior-inferior rim of the glenoid, possible impaction without loose fragment. Note the separated anterior-inferior labrum

Feedback protocol and image evaluation

The feedback team consisted of our two most experienced musculoskeletal radiologists (MvK: 18 years of experience and MvG: 7 years of experience, respectively) and our 2

most experienced orthopaedic shoulder surgeons (CvL: 10 years of experience and PK: 8 years of experience, respectively). Both radiologists are trained in musculoskeletal imaging and evaluate approximately 125 MRA's per year. The orthopaedic surgeons perform approximately 100 diagnostic or therapeutic shoulder arthroscopies each year.

Before assessment of the 20 new MRA's, 45 surgically confirmed MRA's of a former study of ours [7] were used in three feedback sessions, to enhance personal feedback in a short period of time and fine-tune the agreement about lesion definition interpretation between the radiologists and orthopaedic shoulder surgeons. During surgery and MRA evaluation in our former study, the same 7 instability related lesions were systematically inspected by the same radiologists / orthopaedic surgeons and with the same scoring list used in our present study.

Figure 5: Example of disagreement about a superior labrum anterior-to-posterior lesion (oblique coronal orientation, T1 weighted): Surgical outcome: no SLAP lesion. MRA assessment: ranging from normal to SLAP lesion. Arrow on MRA: irregular contrast in biceps anchor could be suggestive for a SLAP lesion

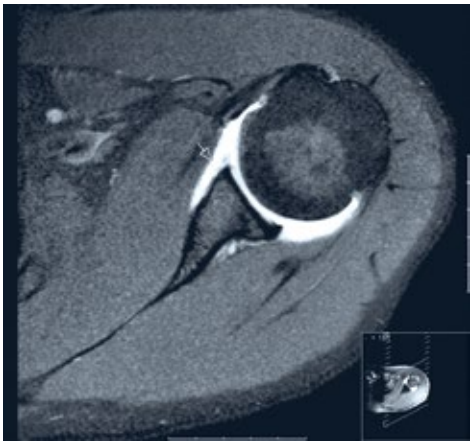

Figure 6: Example of disagreement about a joint capsule lesion (axial orientation, T1 weighted): Surgical outcome: anterior stripping of the anterior capsulo-ligamentous complex from the anterior rim of the glenoid. MRA assessment: ranging from capsule redundancy to anterior stripping. Arrow on MRA: slight suggestion of anterior capsulo-ligamentous complex disruption from the anterior rim of the glenoid

In feedback session 1 the differences in outcome, regarding greater humeral tuberosity fractures, cuff, Hill-Sachs and bony Bankart lesions, between the MRA assessments of the radiologists and surgical findings of the orthopaedic surgeons were thoroughly discussed by the 2 disciplines. During this session the 2 disciplines scrolled through the MRA sequences together and decided per discrepancy in outcome, whether fine-tuning of definition interpretation between the 2 disciplines was in order or alteration of outcome of the MRA report or surgical scoring list would be appropriate. In feedback session 2 the same was done for classic Bankart, SLAP and joint capsule lesions. After these 2 sessions the radiologists and the orthopaedic surgeons agreed that the radiologists overestimated the amount of SLAP lesions, slightly overestimated the number of cuff lesions and underestimated the amount of classic Bankart lesions. The orthopaedic surgeons were, however, asked to carefully score the cuff lesions and Hill-Sachs lesions as they overlooked a few that were clearly seen on MRA. Before session three, the radiologists were asked to assess 10 old surgery confirmed MRA's again, to verify whether session 1 and 2 improved the diagnostic performance of the radiologists. In feedback session 3 all differences in outcome between the radiologists and the surgical findings of the seven different lesion types were evaluated. During the timeframe of these 3 feedback sessions, the radiologists were asked to attend 2 surgical shoulder stabilization operations of one of our most experienced orthopaedic shoulder surgeons (PK), to directly verify their own MRA report during surgery and to improve three-dimensional orientation, regarding the complex shoulder anatomy.

After attendance of the stabilizing surgeries our two most experienced musculoskeletal radiologists individually assessed the first 10 new MRA's, according to the same seven-item scoring list as was used during surgery, to evaluate the reproducibility and accuracy of MRA of the present prospective diagnostic study. To optimize the diagnostic performance of the radiologist even further, all differences in outcome between the radiologists and surgical findings of the first 10 new MRA's were evaluated in feedback session 4, after which the last 10 new MRA's were assessed by the radiologists. During the assessment of the 20 new MRA's, the radiologists were blinded to patient data, surgical report and MRA findings of other radiologists. The radiologists and orthopaedic surgeon were, however, provided with a list of definitions of all lesion types to assure uniformity.

Statistical analysis

Statistical evaluations were carried out using 20.0 SPSS software. MRA reproducibility of the radiologist was expressed using kappa (k) coefficients and absolute percentage of agreement per different lesion type and overall per 7 lesion types. Kappa values can be categorised as poor (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80) and almost perfect (0.81–1.00) agreement. [42] MRA accuracy, as confirmed by the surgical scoring list, was expressed in sensitivity, specificity and % of correct

diagnosis values, per lesion type and overall per 7 lesion types. Differences in the percentage of correct diagnosis between the two radiologists were tested for significance, using the McNemar test for paired proportions, per lesion type and overall per seven lesion types. [43] The 7 assessed lesion types were dichotomized to construct 2x2 tables to facilitate calculations. No adjustment for multiple testing was made. A p-value of ≤ 0.05 was considered statistically significant.

Results

20 MRA's, 17 men and 3 women, met all inclusion criteria. No sequence was missed during imaging. The mean age at MRA-time was 30.1 years (range 16 to 56). The mean time from MRA to surgery was 3.8 months (range 1 to 17). Table 2 summarizes the frequencies of pathologies seen during surgery or on MRA, according to the standard seven-item scoring list. All data of the two radiologists and orthopaedic surgeon are complete.

Per different lesion type the reproducibility of the radiologists in our present study ranged from k = 0.44 to k = 1.00 (Table 3). Agreement was "moderate" in case of cuff, joint capsule and bony Bankart lesions, "substantial" in case of SLAP lesions and "almost perfect" in case of greater humeral tuberosity fractures, Hill-Sachs and classic Bankart lesions. Per 7 lesion types the overall kappa of our two most experienced musculoskeletal radiologists was "almost perfect" (k = 0.81) and exceedingly higher than the "moderate" overall agreement (k = 0.48) of our former study. The overall absolute percentage of agreement of the two radiologists (90.7%) was also dramatically increased in comparison with the results of our former study (78.2%).

The percentage of correct diagnosis was very high per different lesion type and ranged from 80.0% to 100.0% for both radiologists, in the present study, with the exception of joint capsule lesions (Table 4). Radiologist 1 was only accurate in 40.0% and radiologist 2 in 45.0%. Both radiologists were even less accurate, in diagnosing joint capsule lesions, than in our former study (42.1% and 47.4%). Per 7 lesion types, the overall sensitivity of radiologist 1 (87.8% versus 45.9%), the overall sensitivity and specificity of radiologist 2 (79.6% versus 63.5% and 85.7% versus 80.1%, respectively) were increased in comparison with the results of our former study. Furthermore, the overall accuracy of both radiologists was also exceedingly higher in the present study (85.7% and 83.6%) when compared to our former study (74.4% and 74.8%). There were no significant differences in percentage of correct diagnosis between the two radiologist, per lesion and per 7 lesion types in the present study (Table 5).

Table 2: Frequency of pathologies seen during surgery or on MRA (present study N=20)

Items standard scoring list		*Surgery*	*R1*	*R2*
Cuff lesion:				
	Not present	18	18	18
	Partial tear	1	1	1
	Full thickness tear	1	1	1
Humeral head lesion:				
	Not present	5	4	4
	Degeneration	0	3	3
	Hill-Sachs	15	13	13
Anterior inferior glenoid lesion:				
	Not present	19	18	15
	Degeneration	0	0	0
	Defect/Bony Bankart	1	2	5
Greater humeral tuberosity fractures:				
	Not present	20	20	20
	Present	0	0	0
Anterior inferior labrum lesion:				
	Not present	1	0	1
	Degeneration/Fraying	0	0	0
	Classic Bankart	19	20	19
SLAP lesion:				
	Not present	18	16	18
	Present	2	4	2
Joint capsule lesion:				
	Not present	10	4	9
	Capsule redundancy	5	14	8
	Anterior stripping/Tear	5	2	3
Quality of MRA:				
	Good	20	11	13
	Moderate	0	9	7
	Poor	0	0	0

R1: radiologist 1, R2: radiologist 2, MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view

Table 3: Reproducibility of MRA diagnosis between the individual radiologists of the present study (N=20) and our former study (N=51), per lesion type and overall per 7 lesion types (N=20: 140 pair-wise ratings, N=51: 357 pair-wise ratings)

	Radiologist 1 / Radiologist 2 (present study)	***Radiologist 1 / Radiologist 2 (former study)***
Cuff lesion‡	0.44 (90.0%)	0.70 (88.2%)
Humeral head lesion‡	1.00 (100%)	0.60 (80.4%)
Anterior inferior glenoid lesion‡	0.50 (85%)	0.37 (94.1%)
Greater humeral tuberosity fractures‡	1.00 (100%)	1.00 (100.0%)
Anterior inferior labrum lesion‡	NA (95%)	0.23 (52.9%)
SLAP lesion‡	0.62 (90.0%)	0.05 (58.8%)
Joint capsule lesion‡	0.47 (75.0%)	0.34 (72.5%)
Overall absolute percentage of agreement	**90.7%**	**78.2%**
Overall kappa (95% CI)	**0.81 (0.71-0.91)**	**0.48 (0.38-0.57)**

NA: not applicable (calculation of kappa value not possible due to empty cells in 2x2 table), MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view, ‡: κ-coefficients and absolute percentage of agreement

Table 4: Accuracy (% Se, SP or Acc) of MRA diagnoses of the individual radiologists of the present study (N=20) and our former study (N=38), per lesion type and overall per 7 lesion types (N=20: 140 pair-wise ratings, N=38: 266 pair-wise ratings)

	Radiologist 1 (present study)			***Radiologist 2 (present study)***		
	Sens%	Spec%	Acc%	Sens%	Spec%	Acc%
Cuff lesion	100.0	100.0	100.0	50.0	94.4	90.0
Humeral head lesion	80.0	80.0	80.0	80.0	80.0	80.0
Anterior inferior glenoid lesion	100.0	94.7	95.0	100.0	78.9	80.0
Greater humeral tuberosity fractures	100.0	100.0	100.0	100.0	100.0	100.0
Anterior inferior labrum lesion	100.0	0.0	95.0	100.0	100.0	100.0
SLAP lesion	100.0	88.9	90.0	50.0	94.4	90.0
Joint capsule lesion	70.0	10.0	40.0	50.0	40.0	45.0
Overall Se, Sp and Acc	**87.8**	**84.6**	**85.7**	**79.6**	**85.7**	**83.6**

Table 4: Continued

	Radiologist 1 (former study)			*Radiologist 2 (former study)*		
	Sens%	Spec%	Acc%	Sens%	Spec%	Acc%
Cuff lesion	80.0	90.9	89.5	80.0	78.8	78.9
Humeral head lesion	68.2	56.3	63.2	77.7	43.8	63.2
Anterior inferior glenoid lesion	0.0	94.6	92.1	100.0	97.3	97.4
Greater humeral tuberosity fractures	100.0	100.0	100.0	100.0	100.0	100.0
Anterior inferior labrum lesion	41.9	85.7	50.0	80.6	42.9	73.7
SLAP lesion	50.0	86.1	84.2	50.0	63.9	63.2
Joint capsule lesion	25.0	71.4	42.1	25.0	85.7	47.4
Overall Se, Sp and Acc	**45.9**	**87.8**	**74.4**	**63.5**	**80.1**	**74.8**

MRA: magnetic resonance arthrography, SLAP: superior labral anterior-to-posterior lesion, ABER: abduction external rotation view, Sens: sensitivity percentage, Spec: specificity percentage, Acc: percentage of correctly diagnosed lesions calculated as (true-positive + true-negative) / total patients in 2x2 table

Table 5: Differences in percentages correctly diagnosed, between the individual radiologists in the present study (N=20) and our former study (N=38), are tested using the McNemar test for paired proportions. P-values per lesion type and overall per 7 lesion types (N=20: 140 pair-wise ratings, N=38: 266 pair-wise ratings)

	Radiologist 1 / Radiologist 2 (present study)	*Radiologist 1/ Radiologist 2 (former study)*
Cuff lesion	NA	0.125
Humeral head lesion	1.000	1.000
Anterior inferior glenoid lesion	0.250	0.500
Greater humeral tuberosity fractures	NA	NA
Anterior inferior labrum lesion	NA	0.035*
SLAP lesion	1.000	0.057
Joint capsule lesion	1.000	0.727
Overall, per 7 lesion types	**0.581**	**1.000**
Which radiologist scores best?	**R1**	**R2**

*NA: not applicable, calculation not possible due to empty cells in 2x2 table, MRA: magnetic resonance arthrography, SLAP: superior labrum anterior-to-posterior lesion, ABER: abduction external rotation view *: significant difference at p-level ≤ 0.05*

Discussion

In the present study we were able to support our hypothesis that the implementation of our feedback protocol improves the reproducibility and accuracy of high-field MRA by experienced musculoskeletal radiologists in patients with traumatic anterior shoulder instability.

After implementation of our feedback protocol, the overall absolute percentage of agreement was 90.7%, meaning that the radiologist agreed 127 times out of 140 cases (7 lesion x 20 MRA's = 140 pair wise ratings) about the presence and severity of 7 different lesion types assessed by MRA. The overall reproducibility of the radiologist was "almost perfect" with an overall kappa of 0.81 (CI: 071-091) for instability related lesions. Our result is situated in the high end of the range (Kappa: -0.03 to 0.84) mentioned in previous literature [3,7,6,9,21,24,25,30,32] and is exceedingly better than the results of our former prospective diagnostic study of 51 MRA's with ABER, where our most experienced musculoskeletal radiologists only reached a "moderate" overall kappa of 0.48 in the assessment of 7 instability related shoulder lesions on high-field MRA. [7]

The same pattern occurred with our accuracy results. After execution of our feedback protocol our two radiologists reached high overall sensitivity (87.8% and 79.6%, respectively), high overall specificity (84.6% and 85.7%) and high overall accuracy (85.7% and 83.6%), regarding the 7 instability related lesions. Our results are perfectly in line with the high sensitivity (79%-100%) and specificity (85%-100%) values mentioned by other authors [2,8,10,13,16-21,25,27,28,30,33-35] and much better than our results before the implementation of the feedback protocol. In our previous prospective diagnostic study, the radiologists were only able to reach a rather low overall accuracy (74.4% and 74.8%, respectively) after assessment of 38 shoulder MRA's with ABER for 7 instability related lesions. [7]

It seems that our feedback protocol was indeed able to improve both reproducibility and accuracy of high-field MRA assessments by experienced musculoskeletal radiologists, through enhancement of personal feedback after surgical stabilization and fine-tuning of agreement about lesion definition interpretation between the radiologists and between radiologists and orthopaedic surgeons. The improvement of both values is equally important because a systematic accurate MRA diagnosis of each separate radiologist is only achievable through high reproducibility. Improvement of accuracy will result in better planning of stabilizing shoulder surgery and ultimately result in better healthcare. Although there is no literature to confirm our results, we feel that personal feedback should not only be part of normal radiological training, but discussion about discrepancies (between imaging results and surgical outcome) should be incorporated in normal daily clinical practice of all radiologic and orthopaedic departments.

Our overall accuracy could have been even higher, if the accuracy results of the capsule joint lesions had been any better. Our experienced musculoskeletal radiologists

were accurate in 80-100% for 6 instability related lesion types, but the results for joint capsule lesions were rather disappointing with an accuracy of 40.0% and 45.0%, respectively. Although arthroscopy enables the surgeon to fully examine the anterior part of the joint and the surgeon should therefore be theoretically able to detect anterior capsule lesions, it was discussed that especially redundant capsules were difficult to quantify during arthroscopy. The anterior capsule can not be fully seen through the posterior portal. The camera has to be changed to one of the anterior portals. At the time of the study this was not routinely done by the surgeon. Our reasoning is supported by Dinter et al. [20] They assessed 29 shoulder MRA's for instability and stated that anterior joint capsule redundancy was not unambiguously quantifiable by arthroscopy and that the amount of capsule distension could only be visually estimated. If this is the case, our radiologists had no good chance of ever achieving high accuracy for this particular lesion. If so, it can be argued whether arthroscopy is suitable as a reference standard for anterior joint capsule lesions.

Although we did our utmost to conduct a methodological sound study, our study still suffered from a few limitations. Only patients who underwent surgery were included and the decision to perform surgery was based on clinical examination and MRA findings. Therefore, (partial) verification bias might have been introduced. Furthermore, shoulder arthroscopy is a surgeon-dependant method, although it is the best reference standard available. At last, we used a single surgeon to determine the reference standard (surgical scoring list). It is mentioned in previous literature [21] that the arthroscopic reference standard should be accomplished in consensus by 2 surgeons, to address the problem of interobserver variability among orthopaedic surgeons. The orthopaedic surgeon was however very experienced in stabilizing shoulder surgery and inter-observer variability was especially addressed in our feedback protocol. We therefore do not think this was of influence on our results. The strength of our study is, that it is the first prospective diagnostic study evaluating the influence of a feedback protocol on the diagnostic reproducibility and accuracy of MRA by experienced musculoskeletal radiologists.

In conclusion we state that the results of our study show, that our feedback protocol improves the diagnostic performance of high-field MRA in patients with instability related shoulder lesions. It addresses disappointing diagnostic performance values of MRA with a solution. Enhancing personal feedback after stabilizing surgery, enables radiologists and orthopaedic surgeons to exchange knowledge, to discuss discrepancies and to fine-tune their definition interpretation agreement. Our feedback protocol increases the congruency in professional development between radiologists and orthopaedic surgeons through cooperation and communication, leading to more accurate diagnosis of instability related lesions on MRA, leading to better surgery planning and ultimately to better healthcare. Personal feedback should not only be part of normal radiological training, but discussion about discrepancies should be incorporated in normal daily clinical practice.

References

1. Bencardino J, Gyftopoulos S, Palmer W. Imaging in Anterior Glenohumeral Instability. Radiology 2013; 269(2):323-337.
2. Drescher, R, von Rothenburg T, Ludwig J, Köster O, Schmid G. Direct MR-arthrography of the shoulder with maximum capsular distension for surgical planning. Fortschr Röntgenstr 2004; 176(10):1469-1474.
3. van Grinsven S, Kesselring FO, van Wassenaer-van Hall HN, Lindeboom R, Lucas C, van Loon CJ. MR arthrography of traumatic anterior shoulder lesions showed modest reproducibility and accuracy when evaluated under clinical circumstances. Arch Orthop Trauma Surg 2007;127(1): 11-17.
4. van Kampen DA, van den Berg T, van der Woude HJ, Castelein RM, Terwee CB, Willems WJ. Diagnostic value of patient characteristics, history, and six clinical tests for traumatic anterior shoulder instability. J Shoulder Elbow Surg 2013;22:1310-19.
5. Acid S, Le Corroller T, Aswad R, Pauly V, Champsaur P. Preoperative imaging of anterior shoulder instability: diagnostic effectiveness of MDCT arthrography and comparison with MR arthrography and arthroscopy. AJR Am J Roentgenol 2012;198(3):661-667.
6. van der Veen HC, Collins JP, Rijk PC. Value of magnetic resonance arthrography in post-traumatic anterior shoulder instability prior to arthroscopy: a prospective evaluation of MRA versus arthroscopy. Arch Orthop Trauma Surg 2012;132(3):371-375.
7. van Grinsven S, Hagenmaier F, van Loon CJ, van Gorp MJ, van Kints MJ, van Kampen A. Does the experience level of the radiologist, assessment in consensus, or the addition of the abduction and external rotation view improve the diagnostic reproducibility and accuracy of MRA of the shoulder?. Clinical Radiology 2014; 69:1157-1164.
8. Khedr SA, Kassem HM, Azab MA. Comparative study of direct MR arthrography and CT arthrography with arthroscopic correlation in preoperative evaluation of anterior shoulder instability. Egypt J Radiol Nucl Med 2013;44:817-827.
9. Simao MN, Nogueira-Barbosa MH, Muglia VF, Barbieri CH. Anterior shoulder instability: correlation between magnetic resonance arthrography, ultrasound arthrography and intraoperative findings. Ultrasound Med Biol 2012;38(4):551-560.
10. Smith TO, Drew BT, Toms AP. A meta-analysis of the diagnostic test accuracy of MRA and MRI for the detection of glenoid labral injury. Arch Orthop Trauma Surg 2012;132(7):905-919.
11. van der Woude HJ, Vanhoenacker FM. MR Arthrography in Glenohumeral Instability. JBR–BTR 2007; 90:377-383.
12. Bak K, Wiesler ER, Poehling GG. Consensus statement on shoulder instability. Arthroscopy 2010;26(2):249-255.
13. Oh JH, Kim JY, Choi JA, Kim WS. Effectiveness of multidetector computed tomography arthrography for the diagnosis of shoulder pathology: comparison with magnetic resonance imaging with arthroscopic correlation. J Shoulder Elbow Surg 2010;19(1):14-20.
14. Pandya NK, Colton A, Webner D, Sennett B, Huffman GR. Physical examination and magnetic resonance imaging in the diagnosis of superior labrum anterior-posterior lesions of the shoulder: a sensitivity analysis. Arthroscopy 2008;24(3):311-317.

15. Amin MF, Youssef AO. The diagnostic value of magnetic resonance arthrography of the shoulder in detection and grading of SLAP lesions: Comparison with arthroscopic findings. Eur J Radiol 2012;81(9): 2343-2347.
16. Genovese E, Spano E, Castagna A, Leonardi A, Angeretti MG, Callegari L, et al. MR-arthrography in superior instability of the shoulder: correlation with arthroscopy. Radiol med 2013;118:1022–1033.
17. Iqbal HJ, Rani S, Mahmood A, Brownson P, Aniq H. Diagnostic value of MR arthrogram in SLAP lesions of the shoulder. Surgeon 2010;8(6):303-309.
18. de Jesus JO, Parker L, Frangos AJ, Nazarian LN. Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. AJR Am J Roentgenol 2009;192(6):1701-1707.
19. Tian CY, Cui GQ, Zheng ZZ, Ren AH. The added value of ABER position for the detection and classification of anteroinferior labroligamentous lesions in MR arthrography of the shoulder. Eur J Radiol 2013; 82(4):651-657.
20. Dinter DJ, Martetschläger F, Büsing KA, Schönberg SO, Scharf HP, Lehmann LJ. Shoulder Injuries in Overhead Athletes: Utility of MR Arthrography. Sportverl Sportschad 2008;22:146–152.
21. Holzapfel K, Waldt S, Bruegel M, et al. Inter- and intraobserver variability of MR arthrography in the detection and classification of superior labral anterior posterior (SLAP) lesions: evaluation in 78 cases with arthroscopic correlation. Eur Radiol 2010;20(3):666-673.
22. Jonas SC, Walton MJ, Sarangi PP. Is MRA an unnecessary expense in the management of a clinically unstable shoulder? Acta Orthop 2012;83(2):1-4.
23. Reuss BL, Schwartzberg R, Zlatkin MB, Cooperman A, Dixon JR. Magnetic resonance imaging accuracy for the diagnosis of superior labrum anterior-posterior lesions in the community setting: eighty-three arthroscopically confirmed cases. J Shoulder Elbow Surg 2006;15(5):580-585.
24. Schreinemachers SA, van der Hulst VP, Willems WJ, Bipat S, van der Woude HJ. Detection of partial-thickness supraspinatus tendon tears: is a single direct MR arthrography series in ABER position as accurate as conventional MR arthrography? Skeletal Radiol 2009;38(10):967-975.
25. Schreinemachers SA, van der Hulst VP, Jaap Willems W, Bipat S, van der Woude HJ. Is a single direct MR arthrography series in ABER position as accurate in detecting anteroinferior labroligamentous lesions as conventional MR arthography? Skeletal Radiol 2009;38(7):675-683.
26. Stetson WB, Phillips T, Deutsch A. The use of magnetic resonance arthrography to detect partial-thickness rotator cuff tears. J Bone Joint Surg Am 87 Suppl 2005;2:81-88.
27. Toyoda H, Ito Y, Tomo H, Nakao Y, Koike T, Takaoka K. Evaluation of rotator cuff tears with magnetic resonance arthrography. Clin Orthop Relat Res 2005;439:109-115.
28. Theodoropoulos JS, Andreisek G, Harvey EJ, Wolin P. Magnetic resonance imaging and magnetic resonance arthrography of the shoulder: dependence on the level of training of the performing radiologist for diagnostic accuracy. Skeletal Radiol 2010;39(7):661-667.
29. Ito Y, Sakai T, Tomo H, et al. Computerized assessment of Bankart lesions under tension with magnetic resonance arthrography. J Shoulder Elbow Surg 2005;14(3):247-251.
30. Jung JY, Jee WH, Chun HJ, Ahn MI, Kim YS. Magnetic resonance arthrography including ABER view in diagnosing partial-thickness tears of the rotator cuff: accuracy, and inter- and intra-observer agreements. Acta Radiol 2010;51(2):194-201.

31. Song HT, Huh YM, Kim S, Lee SA, Kim SJ, Shin KH, et al. Anterior-inferior labral lesions of recurrent shoulder dislocation evaluated by MR arthrography in an adduction internal rotation (ADIR) position. J Magn Reson Imaging 2006;23(1):29-35.
32. Halma JJ, Eshuis R, Krebbers YM, Weits T, de Gast A. Interdisciplinary inter-observer agreement and accuracy of MR imaging of the shoulder with arthroscopic correlation. Arch Orthop Trauma Surg 2012;132(3):311-320.
33. Applegate GR, Hewitt M, Snyder SJ, Watson E, Kwak S, Resnick D. Chronic labral tears: value of magnetic resonance arthrography in evaluating the glenoid labrum and labral-bicipital complex. Arthroscopy 2004;20(9):959-963.
34. Hayes ML, Collins MS, Morgan JA, Wenger DE, Dahm DL. Efficacy of diagnostic magnetic resonance imaging for articular cartilage lesions of the glenohumeral joint in patients with instability. Skeletal Radiol 2010;39(12):1199-1204.
35 Jana M, Srivastava DN, Sharma R, Gamanagatti S, Nag HL, Mittal R, et al. Magnetic resonance arthrography for assessing severity of glenohumeral labroligamentous lesions. J Orthop Surg 2012;20(2):230-235.
36. Blum A, Coudane H, Molé D. Gleno-humeral instabilities. Eur Radiol 2000;10:63-82.
37. Matsen FA, Rockwood CA. Glenohumeral instability. The Shoulder. Matsen FA, Rockwood CA, editors. WB Saunder Company, Philadelphia 1998; 611-745.
38. Sailer J, Imhof H. Schultergelenkinstabilität. Radiologe 2004;44:578-590.
39. Sanders TG, Morisson WB, Miller MD. Imaging Techniques for the Evaluation of Glenohumeral instability. Am J Sports Med 2000;28:414-434.
40. Snyder SJ, Karzel RP, Del Pizzo W, Ferkel RD, Friedman MJ. SLAP lesions of the shoulder. Arthroscopy 1990;6:274-279.
41. Woertler K, Waldt S. MR imaging in sports-related glenohumeral instability. Eur Radiol 2006;16: 2622-2636.
42. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psycholog Bull 1968;70(4):213-220.
43. Altman DG. Statistics in medical journals: developments in the 1980s. Stat Med 1991;10(12):1897-1913.

Chapter | 8

Summary and conclusions

The research questions addressed in this thesis are answered per chapter. Conclusions are drawn and the implications for the daily clinical practice of radiologists and orthopaedic shoulder surgeons are outlined.

Summary and conclusion per chapter

Research question 1: At what point should a general practitioner (GP) refer a patient with anterior shoulder dislocation or complaints after reduced anterior shoulder dislocation to an orthopaedic shoulder surgeon to avoid further damage / complaints?

Chapter 2 offers the GP a practical guideline to the addressed research question. A GP should always refer a patient with suspected traumatic anterior shoulder dislocation to the emergency room (ER) of a hospital, as fractures must be ruled out with scapulolateral and anteroposterior view radiographs.

Due to the high recurrence rate (up to 95%) of traumatic anterior shoulder dislocation in young and active patients, a swift transition to arthroscopic anatomic stabilizing shoulder surgery is advocated. Early referral by the GP to an orthopaedic surgeon, in case of persistent instability, prevents further damage through recurrent dislocation.

In view of the surgical options ranging from rotator cuff repair to reversed total shoulder prosthesis a prompt referral for additional diagnostics can be very useful, in case of older patients, particularly when progress of range of motion and muscle strength rehabilitation is limited.

Instability caused by hyperlaxity can be determined by the GP by means of the sulcus sign test, anterior apprehension test and Beighton scale. Due to the fact that in shoulder hyperlaxity a cuff injury or intra-articular damage are not very likely and the result of surgery is very moderate, referral is here only useful if a conservative stabilizing physiotherapeutic program of 1 year is not successful.

In conclusion: Instability due to hyperlaxity is a poor indication for referral. A GP should, however, always refer to the ER of a hospital in case of acute traumatic anterior shoulder dislocation. Referral is also indicated when progress of rehabilitation is limited.

Research question 2: What is the reproducibility and accuracy of an 1.5 Tesla magnetic resonance arthrography (MRA) of patients with traumatic anterior shoulder instability (TASI) under conditions resembling the daily clinical practice of radiologists?

Research question 3: What is the influence of the experience level of the interpreting radiologist on the diagnostic reproducibility and accuracy of an 1.5 Tesla MRA of patients with TASI?

We hypothesized, that the reproducibility and accuracy rates of MRA assessments in previous literature may be overestimated, as evaluation is mostly done under ideal

circumstances. In **chapter 3** we examined the reproducibility and accuracy of an 1.5 Tesla MRA of patients with TASI under conditions resembling daily clinical practice. These conditions include the variable observer experience, independent assessment instead of consensus assessment and no exclusion of suboptimal MRA's due to time constraint. Ultimately, 40 of the 61 included MRA's were surgically confirmed. The pathologies assessed by the 2 radiologists were cuff lesion, Hill-Sachs lesion, bony Bankart lesion, Bankart lesion, greater humeral tuberosity fractures, SLAP lesion and joint capsule lesion.

Per 7 lesion types the agreement between the 2 radiologists about the presence of a lesion on MRA was only poor (kappa = 0.21). Furthermore, we found a remarkably low overall accuracy of our 2 radiologists (sensitivity: 50.0%-60.6%, specificity: 87.9%-66.8%), after surgical confirmation. Per seven lesion types the more experienced radiologist, however, correctly diagnosed significant more lesions than his lesser experienced colleague (78.9% versus 65.5%).

In conclusion: Although MRA is the technique of choice for detecting subtle lesions associated with shoulder instability, MRA reports should be regarded with caution by orthopaedic surgeons in daily clinical practice. The experience level of radiologists can affect its reproducibility and accuracy.

Research question 3: What is the influence of the experience level of the interpreting radiologist on the diagnostic reproducibility and accuracy of an 1.5 Tesla MRA of patients with TASI?

Research question 4: What is the influence of assessment in consensus compared to individual assessment of radiologists on the diagnostic performance of an 1.5 Tesla MRA of patients with TASI?

Research question 5: What is the additional value of the abduction and external rotation (ABER) sequence on the diagnostic performance of an 1.5 Tesla MRA of patients with TASI?

In **chapter 4** we presented the results of our evaluation of the influence of observer experience, assessment in consensus and the additional value of ABER on the diagnostic reproducibility and accuracy of an 1.5 Tesla MRA of patients with TASI. We hypothesized, that a higher experience level, consensus assessment and the addition of the ABER-sequence would increase the diagnostic performance. 58 MRA's were ultimately included, in 51 the additional ABER-sequence was performed and 45 MRA's were surgically confirmed.

The overall kappa between the 6 radiologists (R1-R6) and 3 teams (T1-T3) ranged from poor (0.17) to moderate (0.53), sensitivity from 30.6% to 63.5% and specificity from 73.6% to 89.9%. Per seven lesion types the most experienced radiologists (R1-R2: 78.2%) and teams (T1-T2: 81.8%) agreed significantly more about the presence of a lesion on MRA than the lesser experienced radiologists (R3-R4: 70.6% and R5-R6: 70.9%) and teams (T2-T3: 75.1%). The most experienced radiologists (R1: 74.4%, R2: 74.8%, R3:

75.2%) and teams (T1: 75.9%, T2: 78.1%) were, overall, also consistently more accurate after surgical confirmation than the lesser experienced radiologists (R4: 68.8%, R5: 71.8%, R6: 68.4%) and team (T3: 70.8%). Significant differences were found between R1-R4, R3-R4 and T2-T3.

Per seven lesion types the diagnostic performance of consensus assessment was systematically higher than that of individual assessment. Significant differences in agreement were established between T1-T2 and R3-R4. Significant differences in accuracy were established between T2 and R3 or R4.

No overall significant differences in diagnostic performance were found between the radiologists' assessments with and without ABER.

In conclusion: The addition of ABER does not significantly improve the overall diagnostic MRA performance of patients with TASI. The radiologist's experience level and consensus assessment do contribute to a higher reproducibility and accuracy.

Research question 6: Is the diagnostic reproducibility and accuracy of experienced musculoskeletal radiologists assessing 1.5 Tesla MRA's of patients with TASI superior to experienced orthopaedic shoulder surgeons?

In **chapter 5** we compared the diagnostic reproducibility and accuracy of 2 experienced musculoskeletal radiologists with those of 2 experienced orthopaedic shoulder surgeons interpreting instability related shoulder lesions on MRA. We hypothesized, that the diagnostic performance of both radiologists would be superior. 58 MRA's with 7 instability related bony and soft-tissue shoulder lesions were included. These lesions include greater humeral tuberosity fractures, humeral head, glenoid, capsule, labroligamentous-complex and rotator cuff lesions. 45 MRA's were surgically confirmed.

The overall kappa between the 4 observers ranged from fair (0.40) to moderate (0.49). Per 7 lesion types there was no significant difference in percentage of agreement about the presence of a lesion on MRA between the 2 disciplines (78.2% versus 79.0%). The overall sensitivity and specificity of the 4 observers ranged from 45.9% to 76.5% and from 80.1% to 87.8%, respectively. Per 7 lesion types the percentage of correct diagnosis of both orthopaedic surgeons (82.0% and 75.6%) was exceedingly higher after surgical confirmation than that of their radiologic colleagues (74.4% and 74.8%). In case of the most experienced orthopaedic surgeon this difference in accuracy was significant.

In conclusion: Orthopaedic surgeons should rely on their own MRA interpretation if the clinical diagnosis is not confirmed by MRA and disagreement between the radiologist and orthopaedic surgeon exists about the presence of instability related shoulder lesions on MRA.

Research question 4: What is the influence of assessment in consensus compared to individual assessment of radiologists on the diagnostic performance of an 1.5 Tesla MRA of patients with TASI?
Research question 6: Is the diagnostic reproducibility and accuracy of experienced musculoskeletal radiologists assessing 1.5 Tesla MRA's of patients with TASI superior to experienced orthopaedic shoulder surgeons?

In **chapter 6** we compared the diagnostic performance of radiologists with that of orthopaedic shoulder surgeons in 2 large medical centers. Furthermore, we evaluated the influence of consensus assessment compared with individual assessment. We hypothesized, that the MRA assessments of the radiologists would be superior and that consensus assessment would outperform individual assessment. Ultimately, 45 surgery confirmed MRA's of patients with TASI were included.

The overall kappa between the 2 individual radiologists (0.51 and 0.46) and 2 orthopaedic surgeons (0.46 and 0.41) per medical center was moderate. Although the overall percentage of agreement between the radiologists was slightly higher than that between the orthopaedic surgeons in both centers (80.0% versus 77.5% and 75.2% versus 73.7%), there was no significant difference. In each medical center, however, the most experienced orthopaedic surgeon was exceedingly more accurate than both his radiologic colleagues (81.9% versus 72.4% / 74.6% and 76.5% versus 67.3% / 73.7%) per 7 lesion types. In 3 out of 4 times this difference in percentage correct diagnosis was significant.

Overall accuracy improvement through consensus assessment was merely established for the weakest member of each radiologic or orthopaedic team.
In conclusion: Experienced orthopaedic surgeons are more accurate than radiologists in assessing instability related shoulder lesions on MRA. In case of diagnosis disagreement these orthopaedic surgeons should base their treatment decision on their own MRA interpretation. Future research is necessary to determine the added value of consensus assessment.

Research question 7: Does a feedback protocol improve the diagnostic reproducibility and accuracy of experienced musculoskeletal radiologists assessing 1.5 Tesla MRA's of patients with TASI?

In **chapter 7** we presented our feedback protocol in which radiologists systematically received personal feedback of their MRA assessment after surgery, enabling radiologists and orthopaedic surgeons to discuss discrepancies and fine-tune their agreement about lesion definition interpretation. We hypothesized, that the diagnostic reproducibility and accuracy of our 2 most experienced musculoskeletal radiologists assessing 7 different instability related shoulder lesions on MRA would improve after feedback protocol execution. 45 old surgically confirmed MRA's were used to improve the diagnostic performance and 20 new surgically confirmed MRA's were used for verification.

Per 7 lesion types the kappa and percentage of agreement were dramatically increased (k=0.81 versus k=0.48 and 90.7% versus 78.2%, respectively) between the 2 radiologists. The overall sensitivity of radiologist 1 increased from 45.9% to 87.8%, the overall sensitivity of radiologist 2 increased from 63.5% to 79.6% and the overall specificity of radiologist 2 increased from 80.1% to 85.7%. Furthermore, the overall percentage of correct diagnosis of both radiologists was also exceedingly higher (85.7% and 83.6% versus 74.4% and 74.8%) than in our former study.

Conclusion: The implementation of our feedback protocol dramatically improved the reproducibility and accuracy of MRA of patients with TASI by experienced musculoskeletal radiologists.

The implications for daily clinical practice

The impact of MRA assessment of patients with TASI is high because imaging results, such as lesion type, lesion location and lesion severity, may directly influence the treatment decision of the orthopaedic surgeon from a nonsurgical to an arthroscopic or open surgical approach. The treatment success, defined as a painless, functional unrestrained and stable shoulder, depends on the reproducibility and accuracy of the preoperative MRA diagnosis.

The results outlined in this thesis indicate, that - in order to optimize the diagnostic performance in daily clinical practice - instability related MRA's should be assessed by experienced musculoskeletal radiologists or experienced orthopaedic surgeons, as experience level does affect the reproducibility and accuracy.

It also seems fair to suggest, that even experienced musculoskeletal radiologists or experienced orthopaedic surgeons should be stimulated to consult each other in case of hard to diagnose subtle instability related shoulder lesions on MRA, as our results indicate, that diagnostic shortcomings can be complemented by an experienced colleague. Future research is, however, still needed to fully determine the added value of consensus assessment and its cost effectiveness.

Considering the lack of proof, that the ABER sequence actually does improve the diagnostic performance of instability related MRA's, the disadvantages of ABER utilization in general (susceptible to motion artefacts, intolerable due to pain provocation in 8-20% of the patients, interpretation issues related to anatomical changes, time consuming) and cost efficiency requirements in times of paucity, we advocate, that the ABER sequence should not be incorporated in a MRA protocol.

We recommend, however, the incorporation of personal MRA feedback after stabilizing surgery, discussion about discrepancies between imaging results and surgical outcome, and fine-tuning of agreement about lesion definition interpretation between radiologists and orthopaedic surgeons in normal daily clinical practice. It will enhance true professional development, increase diagnostic performance and will ultimately result in an optimal selection of patients eligible for stabilizing surgery and better healthcare.

Personal note

During the research process, outlined in this thesis, I developed the utmost respect for our radiologists and orthopaedic surgeons. The optimalisation of the diagnostic performance of TASI related MRA was not an easy task and sometimes rather confronting. Our specialists, however, never gave in, always had the best interest of our patients in mind and eventually succeeded in reaching the very high end of the reproducibility and accuracy range mentioned in previous literature. By sharing both our disappointments and successes with international literature I am confident, that we did not only improve the healthcare of our patients, but it will inspire others to do the same. I am proud that I could be part of this very interesting journey.

Chapter | 9

Samenvatting en conclusies

De onderzoeksvragen besproken in deze thesis worden per hoofdstuk beantwoord. Conclusies worden getrokken en de implicaties voor de dagelijkse klinische praktijk van radiologen en orthopaedisch chirurgen worden uiteengezet.

Samenvatting en conclusie per hoofdstuk

Onderzoeksvraag 1: Wanneer dient een huisarts een patient met een anterieure schouderluxatie of klachten na reponeren van een anterieure schouderluxatie door te verwijzen naar een schouderspecialist in de tweede lijn om verdere schade / klachten te voorkomen?

In **hoofdstuk 2** wordt de huisarts een praktische handreiking geboden ten aanzien van bovengenoemde onderzoeksvraag. In geval van verdenking op traumatische anterieure schouderluxatie zou een huisarts de patient altijd moeten doorverwijzen naar de spoedeisende hulp (SEH) van een ziekenhuis omdat fracturen uitgesloten moeten worden met behulp van scapulolaterale en anteroposteriore röntgenfoto's.

Bij jonge actieve patiënten wordt, gezien hun hoge recidiefkans (tot 95%) na traumatische anterieure schouderluxatie, bij voorkeur snel tot een scopische anatomische hersteloperatie overgegaan. Vroege doorverwijzing door de huisarts naar de tweede lijn voorkomt verdere schade door reluxaties, bij aanhoudende instabiliteit.

Bij oudere patiënten kan bij beperkte vordering van de revalidatie, met name in het herstel van "range-of-motion" en kracht, een vlotte doorverwijzing voor nadere diagnostiek erg zinvol zijn, gezien de operatiemogelijkheden variërend van cuffhechting tot omgekeerde totaleschouderarthroplastiek.

Instabiliteit door hyperlaxiteit kan door de huisarts zelf worden vastgesteld met behulp van de sulcustest, anterior apprehensiontest en Beighton-schaal. Omdat bij schouderhypermobiliteit een cuffletsel of intra-articularie schade niet waarschijnlijk is en operatieve resultaten zeer matig zijn, is doorverwijzen hier alleen nuttig indien een conservatief stabiliserend fysiotherapeutisch programma van 1 jaar niet helpt.
In conclusie: Instabiliteit door hyperlaxiteit is een slechte indicatie voor doorverwijzing. Bij acute traumatische anterieure schouderluxatie dient de huisarts echter altijd door te verwijzen naar de SEH van een ziekenhuis. Daarnaast is doorverwijzing geïndiceerd bij beperkte vordering van een revalidatieprogramma.

Onderzoeksvraag 2: Wat is de reproduceerbaarheid en accuratesse van een 1.5 Tesla magnetische resonantie arthrografie (MRA) bij patiënten met traumatische anterieure schouder instabiliteit (TASI) onder condities, die vergelijkbaar zijn met de dagelijkse klinische praktijk van radiologen?

Onderzoeksvraag 3: Wat is de invloed van het ervaringsniveau van de beoordelende radioloog op de diagnostische reproduceerbaarheid en accuratesse van een 1.5 Tesla MRA van patiënten met TASI?

We hypothetiseerden dat in de huidige literatuur de reproduceerbaarheid en accuratesse van MRA beoordelingen misschien wordt overschat, omdat evaluatie meestal onder ideale omstandigheden plaats vindt. In **hoofdstuk 3** hebben we de reproduceerbaarheid en accuratesse van een 1.5 Tesla MRA onderzocht van patiënten met TASI onder condities, die vergelijkbaar zijn met de dagelijkse klinische praktijk. Deze condities omvatten het variabele ervaringsniveau van de beoordelaar, individuele beoordeling in plaats van consensus beoordeling en geen exclusie van door tijdsdruk suboptimale MRA's. Uiteindelijk werden 40 van de 61 geïncludeerde MRA's operatief bevestigd. De door de 2 radiologen beoordeelde pathologieën waren cuff laesie, Hill-Sachs laesie, benige Bankart laesie, Bankart laesie, tuberculum majus fractuur van de humerus, SLAP laesie en gewrichtskapsel laesie.

Per 7 laesies was de overeenstemming tussen de 2 radiologen ten aanzien van de aanwezigheid van een laesie op MRA slechts matig (kappa = 0.21). Daarnaast vonden we bij onze radiologen een opvallend lage totaalscore in accuratesse (sensitiviteit: 50.0%-60.6%, specificiteit: 87.9%- 66.8%) na operatieve bevestiging. Per 7 laesies diagnosticeerde de meer ervaren radioloog echter significant meer laesies correct dan de minder ervaren collega (78.9% versus 65.5%).

In conclusie: Ondanks dat MRA de techniek van keuze is bij het detecteren van subtiele laesies, die geassocieerd zijn met schouder instabiliteit, moeten orthopaedisch chirurgen MRA verslagen met enige voorzichtigheid betrachten in de dagelijkse klinische praktijk. Het ervaringsniveau van radiologen kan de reproduceerbaarheid en accuratesse van MRA's beïnvloeden.

Onderzoeksvraag 3: Wat is de invloed van het ervaringsniveau van de beoordelende radioloog op de diagnostische reproduceerbaarheid en accuratesse van een 1.5 Tesla MRA van patiënten met TASI?

Onderzoeksvraag 4: Wat is de invloed van beoordeling in consensus in vergelijking met individuele beoordeling van radiologen op de diagnostische prestatie van een 1.5 Tesla MRA van patiënten met TASI?

Onderzoeksvraag 5: Wat is de additionele waarde van de abductie en exorotatie (ABER) sequentie op de diagnostische prestatie van een 1.5 Tesla MRA van patiënten met TASI?

In **hoofdstuk 4** presenteerden we de resultaten van onze evaluatie van de invloed van ervaringsniveau, beoordeling in consensus en de additionele waarde van ABER op de diagnostische reproduceerbaarheid en accuratesse van een 1.5 Tesla MRA van patiënten met TASI. We hypothetiseerden dat een hoger ervaringsniveau, consensus beoordeling en

additie van de ABER sequentie zou resulteren in een verbeterde diagnostische prestatie. 58 MRA's werden uiteindelijk geïncludeerd, bij 51 was de additionele ABER sequentie uitgevoerd en 45 MRA's werden operatief bevestigd.

Tussen de radiologen (R1-R6) en teams (T1-T3) varieerde de totaalscore van de kappa van slecht (0.17) tot redelijk (0.53), de sensitiviteit van 30.6% tot 63.5% en de specificiteit van 73.6% tot 89.9%. Per 7 laesies was de overeenstemming tussen de meest ervaren radiologen (R1-R2: 78.2%) en teams (T1-T2: 81.8%) ten aanzien van de aanwezigheid van een laesie op MRA significant groter dan tussen de minder ervaren radiologen (R3-R4: 70.6% en R5-R6: 70.9%) en teams (T2-T3: 75.1%). De meest ervaren radiologen (R1: 74.4%, R2: 74.8%, R3: 75.2%) en teams (T1: 75.9%, T2: 78.1%) waren in totaal ook consistent meer accuraat na operatieve bevestiging dan de minder ervaren radiologen (R4: 68.8%, R5: 71.8%, R6: 68.4%) en team (T3: 70.8%). Significante verschillen werden gevonden tussen R1-R4, R3-R4 en T2-T3.

Per zeven laesie typen was de diagnostische prestatie van consensus beoordeling systematisch beter dan die van individuele beoordeling. Significante verschillen in percentage overeenkomst werden bereikt tussen T1-T2 en R3-R4. Significante verschillen in accuratesse werden bereikt tussen T2 en R3 of R4.

In totaal werd er geen significant verschil gevonden in diagnostische prestatie tussen de beoordelingen van radiologen met of zonder ABER.

In conclusie: De additie van ABER verbetert de totale diagnostische prestatie van MRA, bij patiënten met TASI, niet significant. Het ervaringsniveau van de radioloog en boordeling in consensus draagt wel bij tot een hogere reproduceerbaarheid en accuratesse.

Onderzoeksvraag 6: Is de diagnostische reproduceerbaarheid en accuratesse van ervaren musculoskeletale radiologen ten aanzien van de beoordeling van 1.5 Tesla MRA's van patiënten met TASI superieur ten opzichte van ervaren orthopaedisch chirurgen?

In **hoofdstuk 5** werd de diagnostische reproduceerbaarheid en accuratesse van 2 ervaren musculoskeletale radiologen vergeleken met die van 2 ervaren orthopaedisch chirurgen ten aanzien van de interpretatie van instabiliteitgerelateerde schouder laesies op MRA. We hypothetiseerden, dat de diagnostische prestatie van beide radiologen superieur zou zijn. 58 MRA's, met 7 instabiliteitgerelateerde benige en weke-delen schouder laesies werden geïncludeerd. Deze laesies omvatten humerale tuberculum majus fracturen, humeruskop, glenoid, kapsel, labroligamenteus-complex en rotator cuff laesies. 45 MRA's werden operatief bevestigd.

De totaalscore van de kappa varieerde van matig (0.40) tot redelijk (0.49) tussen de 4 beoordelaars. Per 7 laesie typen was er geen significant verschil in percentage overeenkomst tussen de 2 disciplines (78.2% versus 79.0%) ten aanzien van de aanwezigheid van een laesie op MRA. De totaalscore van de sensitiviteit en specificiteit van de 4 beoordelaars varieerde respectievelijk van 45.9% tot 76.5% en van 80.1% tot

87.8%. Het percentage correcte diagnosen van beide orthopaedisch chirurgen (82.0% en 75.6%) was na operatieve bevestiging veel hoger dan dat van hun radiologische collega's (74.4% en 74.8%). Dit verschil in accuratesse was significant, per 7 laesies, in het geval van de meest ervaren orthopaedisch chirurg.
In conclusie: Orthopaedisch chirurgen zouden op hun eigen MRA interpretatie moeten vertrouwen indien de klinische diagnose niet wordt bevestigd door de MRA en de radioloog en orthopaedisch chirurg van mening verschillen ten aanzien van de aanwezigheid van een instabiliteitgerelateerde schouder laesie op de MRA.

Onderzoeksvraag 4: Wat is de invloed van beoordeling in consensus in vergelijking met individuele beoordeling van radiologen op de diagnostische prestatie van een 1.5 Tesla MRA van patiënten met TASI?
Onderzoeksvraag 6: Is de diagnostische reproduceerbaarheid en accuratesse van ervaren musculoskeletale radiologen ten aanzien van de beoordeling van 1.5 Tesla MRA's van patiënten met TASI superieur ten opzichte van ervaren orthopaedisch chirurgen?

In **hoofdstuk 6** werd de diagnostische prestatie van radiologen vergeleken met die van orthopaedisch chirurgen in 2 grote medische centra. Daarnaast evalueerden we de invloed van consensus beoordeling vergeleken met individuele beoordeling. We hypothetiseerden dat de MRA beoordelingen van de radiologen superior zouden zijn en dat consensus beoordeling de individuele beoordeling zou overtreffen. Uiteindelijk werden 45 operatief bevestigde MRA's van patiënten met TASI geïncludeerd.

De totaalscore van de kappa tussen de 2 individuele radiologen (0.51 en 0.46) en 2 orthopaedisch chirurgen (0.46 en 0.41) was redelijk per medisch centrum. Ondanks het feit, dat in beide centra het percentage overeenkomst tussen de radiologen iets hoger was dan tussen de orthopaedisch chirurgen (80.0% versus 77.5% en 75.2% versus 73.7%), was er geen significant verschil per 7 laesies. In elk medisch centrum was de totale accuratesse van de meest ervaren orthopaedisch chirurg echter veel hoger dan van zijn beide radiologische collega's (81.9% versus 72.4% / 74.6% en 76.5% versus 67.3% / 73.7%). In 3 van de 4 keer was het verschil in percentage correcte diagnoses significant.

Verbetering van de totaalscore in accuratesse door beoordeling in consensus werd alleen bereikt voor het zwakste lid van elk radiologisch of orthopaedisch team.
In conclusie: Ervaren orthopaedisch chirurgen zijn meer accuraat dan radiologen in de beoordeling van instabiliteitgerelateerde schouder laesies op MRA. In geval van verschil van mening ten aanzien van de diagnose zouden deze orthopaedisch chirurgen hun behandelbeslissing moeten baseren op hun eigen MRA interpretatie. Verder onderzoek is nodig om de additionele waarden van consensus beoordeling te bepalen.

Onderzoeksvraag 7: Verbetert het feedback protocol de diagnostische reproduceerbaarheid en accuratesse van ervaren musculoskeletale radiologen ten aanzien van de beoordeling van 1.5 Tesla MRA's van patiënten met TASI?

In hoofdstuk 7 presenteerden we ons feedback protocol waarbij radiologen systematisch persoonlijk feedback ontvingen over hun MRA beoordeling na operatie, waardoor radiologen en orthopaedisch chirurgen in staat gesteld werden om verschillen te bediscussiëren en hun overeenstemming ten aanzien van laesie definitie interpretatie fijn te slijpen. We hypothetiseerden, dat de diagnostische reproduceerbaarheid en accuratesse van onze 2 meest ervaren musculoskeletale radiologen zouden verbeteren ten aanzien van de beoordeling van 7 verschillende instabiliteitgerelateerde schouder laesies op MRA na feedback protocol executie. 45 oude door operatie bevestigde MRA's werden gebruikt om de diagnostische prestatie te verbeteren en 20 nieuwe door operatie bevestigde MRA's werden gebruikt ter verificatie.

Per 7 laesie typen verbeterde de kappa en het percentage overeenkomst tussen de 2 radiologen (k=0.81 versus k=0.48 en 90.7% versus 78.2%, respectievelijk) dramatisch. De totaalscore van de sensitiviteit van radioloog 1 verbeterde van 45.9% tot 87.8%, de totale sensitiviteit van radioloog 2 verbeterde van 63.5% tot 79.6% en de totaalscore van de specificiteit van radioloog 2 verbeterde van 80.1% tot 85.7%. Verder was ook de totaalscore van het percentage correcte diagnosen van beide radiologen hoger (85.7% en 83.6% versus 74.4% en 74.8%) dan in ons eerder onderzoek.
In conclusie: De implementatie van ons feedback protocol bewerkstelligde een dramatische verbetering van de reproduceerbaarheid en accuratesse van MRA beoordelingen van patiënten met TASI door ervaren musculoskeletale radiologen.

De implicaties voor de dagelijkse klinische praktijk

De impact van MRA beoordeling van patiënten met TASI is groot omdat het resultaat, zoals het laesie type, de laesie locatie en de ernst van de laesie, de behandelbeslissing van de orthopaedisch chirurg direct kan beïnvloeden van conservatief tot arthroscopisch of open chirurgische benadering. Het behandelsucces, gedefinieerd als een pijnloze, functioneel onbeperkte en stabiele schouder, is afhankelijk van de reproduceerbaarheid en accuratesse van de preoperatieve MRA diagnose.

De resultaten uiteengezet in deze thesis wijzen erop, dat - ten einde de diagnostische prestatie te optimaliseren in de dagelijkse klinische praktijk - instabiliteitgerelateerde MRA's beoordeeld zouden moeten worden door ervaren musculoskeletale radiologen of ervaren orthopaedisch chirurgen, omdat het ervaringsniveau van invloed is op de reproduceerbaarheid en accuratesse.

Het lijkt tevens gerechtvaardigd om voor te stellen, dat zelfs ervaren musculoskeletale radiologen of ervaren orthopaedisch chirurgen gestimuleerd zouden moeten worden om

elkaar te consulteren in geval van moeilijk te diagnosticeren subtiele instabiliteitgerelateerde schouder laesies op MRA, omdat uit onze resultaten blijkt, dat diagnostische tekortkomingen gecomplementeerd kunnen worden door een ervaren collega. Toekomstig onderzoek is echter nog steeds nodig om de toegevoegde waarde van consensus beoordeling en de kosteneffectiviteit ervan volledig vast te stellen.

Gezien het gebrek aan bewijs dat de ABER sequentie daadwerkelijk de diagnostische prestatie van instabiliteitgerelateerde MRA's verbetert, de nadelen van ABER toepassing in het algemeen (gevoelig voor bewegingsartefacten, onverdraaglijk voor 8-20% van de patiënten door pijnprovocatie, interpretatie problemen gerelateerd aan anatomische veranderingen, tijdsconsumerend) en kosteneffectiviteiteisen in tijden van schaarste, bepleiten we dat de ABER sequentie niet opgenomen zou moeten worden in een MRA protocol.

We adviseren echter incorporatie van persoonlijke MRA terugkoppeling na stabiliserende operatie, discussie over discrepanties tussen MRA resultaten en operatieve uitkomst, en het fijn-slijpen van overeenstemming tussen radiologen en orthopaedisch chirurgen ten aanzien van laesie definitie interpretatie in de dagelijkse klinische praktijk. Het verbetert de daadwerkelijke professionele ontwikkeling, verbetert de diagnostische prestatie en zal uiteindelijk resulteren in een optimale selectie van patiënten, die in aanmerking komen voor een stabiliserende operatie en een verbeterde gezondheidszorg.

Persoonlijke notitie

Gedurende het onderzoeksproces, besproken in deze thesis, heb ik een groot respect gekregen voor onze radiologen en orthopaedisch chirurgen. De optimalisatie van de diagnostische prestatie van TASI gerelateerde MRA was geen gemakkelijke opgave en soms tamelijk confronterend. Onze specialisten gaven zich echter nooit gewonnen, hielden altijd het belang van onze patiënten voor ogen en slaagden er uiteindelijk in om het hoge uiteinde van de in de literatuur vermelde spreidingsbreedte van de reproduceerbaarheid en accuratesse te bereiken. Door zowel onze teleurstellingen als onze successen met de internationale literatuur te delen ben ik ervan overtuigd, dat we niet alleen de gezondheidszorg van onze patiënten hebben verbeterd, maar dat ook anderen hierdoor geïnspireerd raken om hetzelfde te bewerkstelligen. Ik ben er trots op dat ik deel heb mogen uitmaken van deze zeer interessante reis.

List of Publications

Curriculum Vitae

List of Publications

Hol A, **van Grinsven S**, Rijnberg W, van Susante J, van Loon C. Variatie in nabehandeling van de primaire totaleheupprothese via de posterolaterale benadering.
Ned Tijdschr Orthopaedie. 2006;13:105-113.

van Grinsven S, Kesselring F, Wassenaer-van Hall H, Lindeboom R, Lucas C, van Loon C. MR arthrography of traumatic anterior shoulder lesions showed modest reproducibility and accuracy when evaluated under clinical circumstances.
Arch Orthop Trauma Surg. 2007;127:11-27 [DOI: 10.1007/s00402-006-0205-7].

van Grinsven S, Holla C, van Cingel R, van Loon C. "Evidence based" nabehandeling van voorstekruisbandreconstructie.
Ned Tijdschr Orthopaedie. 2008;15:70-81.

Hol A, **van Grinsven S**, Lucas C, van Susante J, van Loon C. Partial versus unrestricted weight bearing after an uncemented femoral stem in total hip arthroplasty: recommendation of a concise rehabilitation protocol from a systematic review of the literature.
Arch Orthop Trauma Surg. 2009;130(4)547-555 [DOI: 10.1007/s00402-009-1017-3].

van Grinsven S, van Loon C. Schroef- en cementaugmentatie voor kleine botdefecten bij totaleknieprothese – gewapend beton.
Ned Tijdschr Orthopaedie. 2010;17:9-16.

Schmitz M, **van Grinsven S**, Rijnberg W, van Loon C. Frequente schroefbreuk na gebruik Numelock fixatiestysteem bij openwig-tibiakoposteomie.
Ned Tijdschr Orthopaedie. 2010;17:69-74.

van Grinsven S, van Cingel R, Holla C, van Loon C. Evidence-based rehabilitation following anterior cruciate ligament reconstruction.
Knee Surg Sports Traumatol Arthrosc. 2010;18:1128-1144 [DOI: 10.1007/s00167-009-1027-2].

Brokelman R, Haverkamp D, **van Grinsven S**, van Kampen A, Veth R, van Loon C. Tevredenheid bij patiënten met een totaleknieprothese, kunnen we het meten?
Ned Tijdschr Orthopaedie. 2011;18:114-120.

van der Groes S, **van Grinsven S**, Bisseling P, van Loon C. Totale knieprothese bij varus- en valgusdeformiteit: klinische resultaten en objectieve coronale alignement en laxiteitmeting. Ned Tijdschr Orthopaedie. 2012;19:80-87.

Boons H, Goossen J, **van Grinsven S**, van Susante S, van Loon C. Hemiarthroplasty for Humeral Four-part Fractures for Patients 65 Years and Older, A Randomized Controlled Trial.
Clin Orthop Relat Res. 2012;470(12):3483-3491 [DOI: 10.1007/s11999-012-2531-0].

van Grinsven S, Hagenmaier F, van Loon C, van Gorp M, van Kints M, van Kampen A. (2014) Does the experience level of the radiologist, assessment in consensus, or the addition of the abduction and external rotation view improve the diagnostic reproducibility and accuracy of MRA of the shoulder?
Clin Radiol. 2014;69(11):1157-1164 [DOI: 10.1016/j.crad.2014.07.009].

van Egmond N, **van Grinsven S**, van Loon C, Gaasbeek R, van Kampen A. Better clinical results after closed- compared to open-wedge high tibial osteotomy in patients with medial knee osteoarthritis and varus leg alignment.
Knee Surg Sports Traumatol Arthrosc. 2014 Sep 13 [Epub ahead of print PMID: 25217316].

van Grinsven S, Nijenhuis T, Konings P, van Kampen A, van Loon C. Are radiologists superior to orthopaedic surgeons in diagnosing instability related shoulder lesions on magnetic resonance arthrography? A multicenter reproducibility and accuracy study.
J Shoulder Elbow Surg. 2015;24(9):1405-1412 [DOI: 10.1016/j.jse.2015.05.050].

van Grinsven S, van Loon C, van Gorp M, van Kints M, Konings P, van Kampen A.
A feedback protocol improves the diagnostic performance of MR Arthrography by experienced musculoskeletal radiologists in patients with traumatic anterior shoulder instability.
Eur J Radiol. 2015;84:2242-2249 [DOI: 10.1016/j.ejrad.2015.07.016].

Bisseling P, **van Grinsven S**, Rijnberg W, van Loon C. Single system posterior stabilized revision total knee arthroplasty. A minimum of 2-year follow-up analysis.
Ned Tijdschr Orthopaedie. Accepted.

Hartman-Oomen J, **van Grinsven S**, van Loon C. Re-rupture/failure rate after ACL reconstruction.
Ned Tijdschr Orthopaedie. Accepted.

Curriculum Vitae

Susan van Grinsven was born on July 19th, 1968 in Boxmeer (NB). She graduated at the Elzendaal College in Boxmeer in 1987. From 1988 until 1989 she was trained in management and marketing at the Schoevers Private Training Institution. In 1992 she started her general physiotherapy training at the Fontys University of Applied Sciences in Eindhoven where she received her bachelor's degree in 1996. In the same year she joined the physiotherapy department of the Rijnstate Hospital Velp / Arnhem were she specialized in chronic pulmonary diseases and asthma, the shoulder, neuro developmental treatment and intensive care treatment. In 2003 she started her master of science in clinical epidemiology training at the University of Amsterdam which was completed in 2005. In the same year she accepted her current position at the Rijnstate Hospital Arnhem as clinical epidemiologist for the paramedical department and department of orthopaedics. The vast majority of her activities for the department of orthopaedics consist of supervising the scientific research and its publication of medical interns and residents. During her work for the outpatient shoulder clinic of the Rijnstate Hospital Arnhem the first spark of interest was raised for the diagnostic process of shoulder pathology. This eventually led to the start of the research on this thesis under supervision of dr. C.J.M. van Loon (Rijnstate Hospital Arnhem) and prof. dr. A. van Kampen (Radboud University Medical Center Nijmegen) in 2012.